ANNALS

of the Association of American Geographers

DARTMOUTH COLLEGE
APR 24 1961
LIBRARY

ROBERT S. PLATT
Editor

Volume 51 • Number 1

March 1961

CONTENTS

Walled Cities of the United States ..... Howard J. Nelson

Some Aspects of the Urban Geography of the Chinese Hsien Capital ..... Sen-dou Chang

Missions and Mission Lands of Alta California ..... R. Louis Gentilcore

The Dynamics of Store Trading Areas and Market Equilibrium ..... William Applebaum and Saul B. Cohen

The Marine Terraces of Aruba, Bonaire, and Curaçao, Netherlands Antilles ..... Charles S. Alexander

Centrifugal-Centripetal Forces in International Petroleum Company Regions ..... Robert B. McNee

ANNALS of the Association of American Geographers is published quarterly in March, June, September, and December at the Allen Press, Lawrence, Kansas

Subscription rate: $10.00 per calendar year
Volumes 48 and later: $10.00 per volume; $2.50 per issue
Volumes 42–47 incl.: $5.00 per volume; $1.25 per issue
Volume 41: $4.00 per volume; $1.00 per issue

Address all communications relative to subscriptions, purchase of back numbers, etc., to Association of American Geographers, 1785 Massachusetts Ave., N.W., Washington 6, D.C.

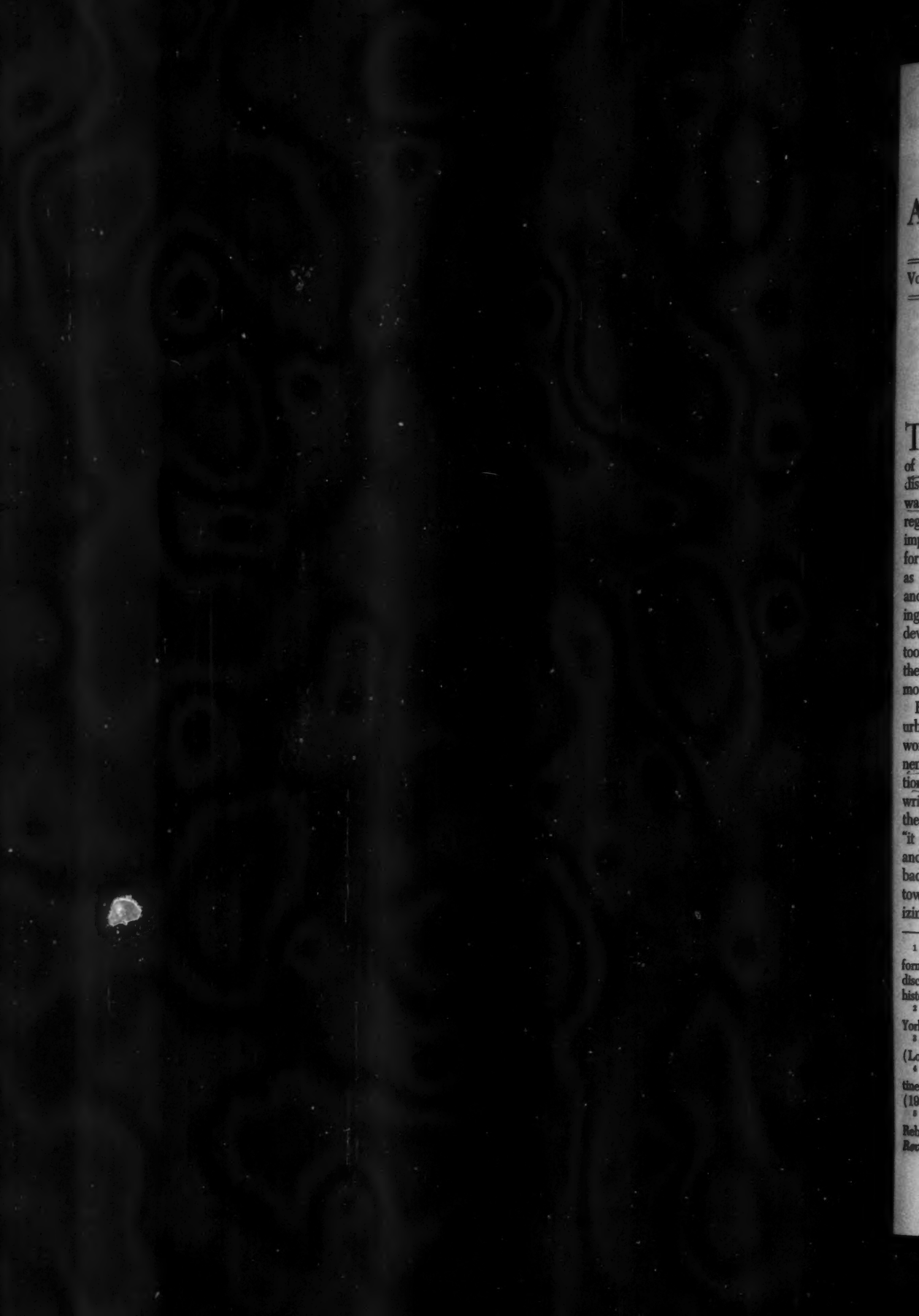
Submit all manuscripts to
Robert S. Platt, *Editor*, Dept. of Geography,
University of Chicago, Chicago 37, Illinois

Submit publications for review and also reviews to
Marion E. Marts, *Review Editor*
Dept. of Geography, University of Washington, Seattle 5

Submit maps for map supplement to
Richard Edes Harrison, *Map Editor*
313 E. 51st Street, New York City 22, N.Y.

Statements and opinions in the ANNALS *are the responsibility of authors and do not necessarily reflect the views of the Association*

Second class postage paid at Lawrence, Kansas, U. S. A.



ANNALS of the
Association of American Geographers

Volume 51 March 1961 Number 1

WALLED CITIES OF THE UNITED STATES[1]

HOWARD J. NELSON
University of California, Los Angeles

THE significance of the "wall" is a common topic in the analysis of cities in most parts of the world. It is perhaps most common in discussions of European cities where many walls are still standing, although even in this region discussion generally focusses on the importance of earlier walls. Lewis Mumford, for example, in his *Culture of Cities* uses this as the theme of his first chapter, "Protection and the Town," in which he discusses in striking detail the significance of the wall in the development of European cities.[2] Dickinson, too, emphasizes the effect of earlier walls on the layout and internal arrangement of many modern European cities.[3]

But city walls are also important topics in urban studies in many other parts of the world. Thomas, working in Africa, prominently mentions the city wall and fortifications in his study of Constantine.[4] Spencer, writing on "Changing Chunking," reports on the evolution of that city's wall until in 1731 "it measured some five miles in circumference and had 17 gates." In 1938 it was reported in bad repair and "most of the gates and bastion towers had long since disappeared."[5] Generalizing, in a later work, he points out that in China "the most common early words for city relate to walls and gates," and that the protective wall was a common part of town building in many sections of the Orient.[6] The student soon comes to think of walls as characteristic features in the development of cities in most parts of the world outside of the United States. Cities in the United States, perhaps logically in terms of their relatively recent development and present unwalled condition, are usually considered from different points of view. While it is common knowledge that Wall Street in New York City follows the line of a former wall, and while the historic walls of an occasional American city are not unknown to individual students of cities, comparative literature on American city walls is lacking.

An important task of the urban geographer is the analysis of the variations in the cities of the world's culture regions. In making such an analysis it is a common practice to use the influence of earlier walls as one of the distinguishing features of growth and internal arrangement of cities in many of the culture regions of the world. Similarly, American cities are usually characterized as examples of urban developments uniformly free from the effect of restricting walls.

The lack of the influence of the wall may be a characteristic feature of cities in the United States. However, European settlers arriving in a land where danger existed from native population and from rival colonists alike, and with a background of familiarity with walled cities in their homelands, quite naturally built walls around a number of their

[1] A Research Grant from the University of California which enabled me to visit the eleven cities discussed in this paper and to walk the lines of their historic walls is gratefully acknowledged.

[2] Lewis Mumford, *The Culture of Cities* (New York: Harcourt Brace, 1938).

[3] Robert E. Dickinson, *The West European City* (London: Kegan Paul, 1951).

[4] Benjamin E. Thomas, "Fortress City of Constantine, Algeria," *The Scientific Monthly*, Vol. 81, No. 3 (1955), pp. 134–35.

[5] Joseph E. Spencer, "Changing Chunking: the Rebuilding of an Old Chinese City," *Geographical Review*, Vol. 29, No. 1 (1939), pp. 56–73.

[6] Spencer, *Asia, East by South* (New York: Wiley, 1954), p. 140.

new settlements. Of the 252 major cities in the United States (those with a population of 50,000 or more in 1950), eleven had walls at an early period in their development. City walls are defined here as man-made barriers defending an essentially civilian settlement. Forts, military posts, or presidios, which occasionally formed the nucleus of a settlement or which stood on the edge of a settlement, are not considered as constituting walled cities. Too, settlements that may have been walled and have disappeared or have not reached major city size have been omitted from this study.[7]

The eleven walled cities include settlements of the major European powers that colonized the continent, as well as several fortified by a later generation of Americans. Three English settlements were walled: Boston, Charleston, and Savannah. Three Dutch towns, too, had walls: New Amsterdam, Albany, and Schenectady. The French walled what was to become Detroit, and the walls were maintained and enlarged by both the English and Americans. New Orleans was walled first by the French and then by the Spaniards, while the wall around St. Louis was built during a period of Spanish sovereignty. In a later period the Mormons in Utah constructed massive walls around both Salt Lake City and Ogden.

It is the purpose of this article to examine the development of the walls of these eleven cities, and with precedents in mind from walled cities in other regions, to assess the effect of these historic walls on the growth and structure of the present cities. For example, were American city walls really comparable to their European counterparts? Where were they located, of what were they built, how long did they last, and were they defensively effective? In Europe today a number of boulevards and other streets follow the lines of former walls. Are there similar examples to be found in American cities? Has the location of former gates aligned streets still present? Are open spaces present or is there land in public use as the result of the location of former fortifications? Are there any other features present in today's urban landscape that reflect the presence of the earlier wall?

[7] Numerous other settlements not large enough to be in our sample group also had walls. They include St. Augustine and Pensacola, Florida, and Provo, Spanish Fork, Fillmore, and Cedar City, Utah. The Canadian cities of Montreal, Quebec, and Halifax also had walls.

BOSTON

The site of Boston, although selected for other reasons—an "excellent spring" and a deep harbor—was ideally located for landward defense. Before the Roxbury Flats (Back Bay) and the South Cove were filled in, only a narrow neck connected the Boston peninsula with the mainland. A colonial visitor, Capt. Nathaniel Uring, assessed the potentialities in 1709:

> The situation of the Town is such that it is capable of being fortified and made as strong as any in Europe, there being only a narrow Isthmus or Neck of Land of about 40 yards broad, which has a Communication with the country, and is so low, that the Spring Tides sometimes washes the Road; which with little change might be fortified and made so strong, that it would be impossible to force it, and no way of coming at it by Land, but over the Neck.[8]

Unknown to Capt. Uring, the citizens of Boston had recognized the defensive quality on the neck and had put it to use in time of crisis. In fact, at intervals over a period of 150 years, the narrow neck was used for defense. On the earliest occasion, in 1631, only a guard was posted: "We began a court of guard upon the neck between Roxbury and Boston, whereon should always be resident an officer and six men."[9] Presumably the watch was soon discontined. Four years later a "gate and style" [*sic*] were ordered built on the neck[10] across the Roxbury road (today Washington Street), and it seems to have remained until the end of the decade, more of a barrier for cows than for men.

However, King Philip's War resulted in the building of the first major barrier against human transgressors. A watch tower was erected

[8] Capt. Nathaniel Uring, *A History of the Voyages and Travels of Capt. Nathaniel Uring* (London: J. Peele, 1726), pp. 109–10.

[9] Journal entry, April 14, 1631. James K. Hosmer, ed., *Winthrop's Journal "History of New England," 1630–1649* (New York: Scribner's Sons, 1908), Vol. I, p. 62.

[10] Agreed at general meeting, January 23, 1635. *Second Report of the Record Commission of the City of Boston; Containing the Boston Records, 1634–1660* (Boston: Rockwell and Churchill, 1887), Vol. II, p. 4.

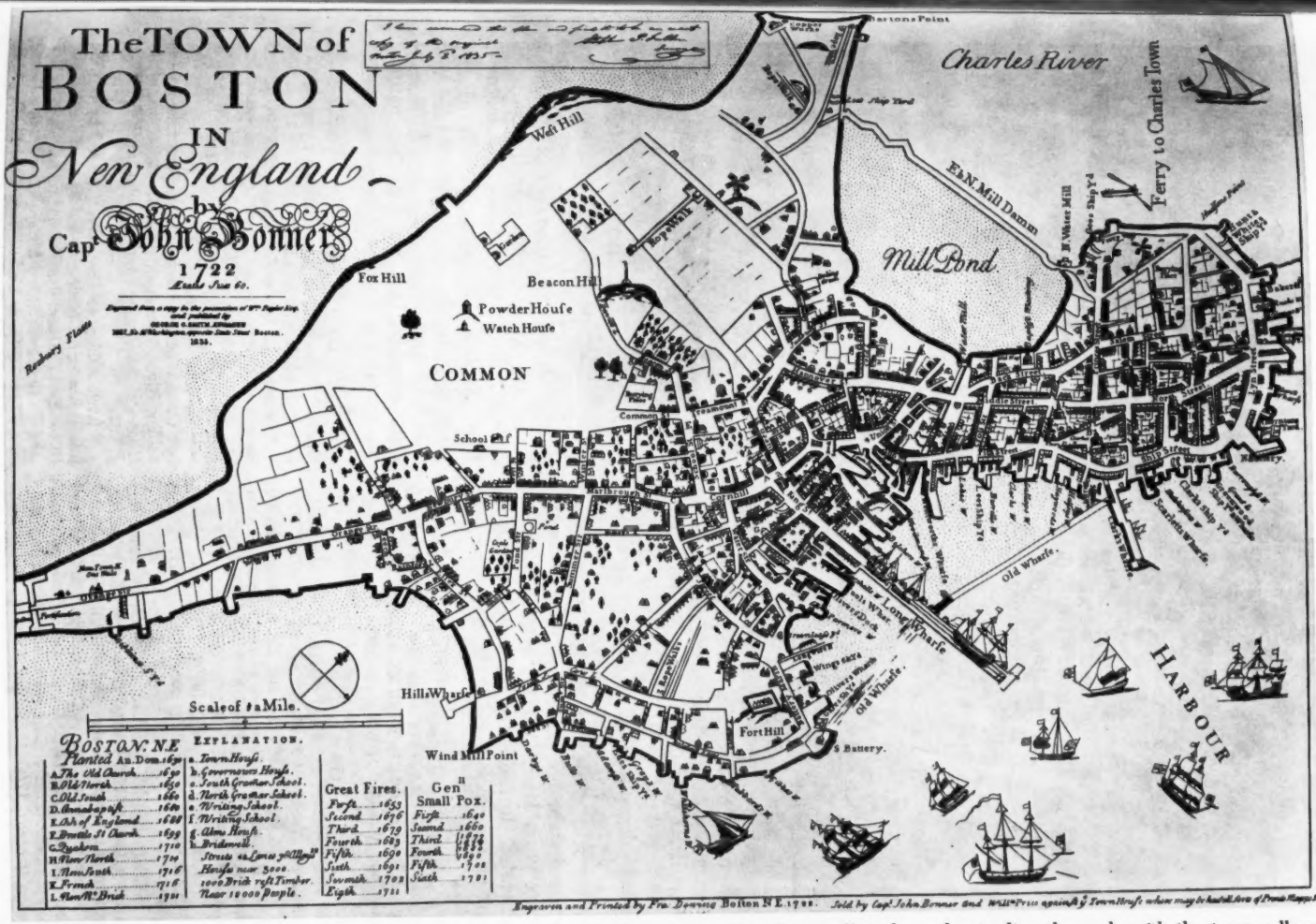


FIG. 1. Boston in 1722 as depicted by the familiar map of Capt. John Bonner. The short wall is shown barricading the neck, with the town gallows indicatd just outside of the gate. (Courtesy of the Bostonian Society.)

on the neck in 1677[11] and two of the eight town constables appointed to "watch every night" were ordered to patrol the neck and "to cause the gate to be shut at night and opened at five in the morning."[12] Fortifications seem to have been in existence in 1700, for in that year they were used as a place of reference on the neck.[13]

Trouble with the French early in the eighteenth century resulted in a more substantial wall. In 1706, 200 pounds, one-fourth of the money appropriated for the defense of the city, was designated for fortifying the neck.[14] Early in 1711, a space "100 yeards in breadth, across from sea to sea . . . where the fortifications stood" (approximately the present Dover Street) was reserved for the town, although other land on the neck was sold.[15] Later in that same year a line of defense was ordered to be made "forthwith" and to be faced with "stone from the bottom of the ditch to the height of the Rampart, and the Parapet to be of Sodworke." It was also requested that the committee "enquire into and recover the stones of the town that have been taken away from off the neck . . . and that the persons who took them away be obliged to return them."[16] The ordering of a ditch lends credence to the later description of Watkins: "Outside the fortifications was a ditch, twenty feet wide, the tide running through and a drawbridge crossing it."[17]

Six years later, in 1717, more money was expended "to finish the brickwork to the extent of the stone-work."[18] In the same year Samuel Sewell reports after a journey from Boston to Roxbury: "I had the pleasure to view the wall of our city and pass in and out of the Gate; the work being closed, postern and all. The Lord keep the city."[19] The line of the wall in this period is illustrated on the Bonner map (Fig. 1) engraved in 1722.

Again, with the threatened French invasion in 1746, it was ordered that "the Ditch or mote be Cleared and Guns provided for the Ports in the Breastwork on the neck, that there be a platform made to Traverse the Guns on. . . ."[20] These fortifications were kept in repair during hostilities with the French and Indians in 1757. They played a part in the siege of Boston during the Revolution when additional lines were built on the neck. They were reportedly destroyed in 1787.[21]

CHARLESTON

The English in Carolina were both fearful of the Indians and close to Spanish Florida, and thus conscious of the need for defense. Their first settlement on the Ashley River boasted a palisade;[22] even so, it was gradually abandoned for the site of Charleston. This site, a low bluff on a narrow peninsula between the Cooper and Ashley rivers, was occupied after 1677 and became the seat of government in 1680. Lord Ashley directed Governor West to lay out a "port Town into regular streets," and thought that "A Pallisado round the Towne with a small ditch is sufficient Fortification against the Indians."[23] Gov. Archdale, arriving in 1695, was also ordered to fortify Charleston, and the work was continued by his successors.[24] An act of the Assembly, Dec. 23, 1703, for "fortifying the remaining part of Charleston," ordered

[11] Vote at general meeting, March 12, 1676/7. *A Report of the Record Commission Containing the Boston Records for 1660 to 1701* (Boston: Rockwell and Churchill, 1881), Vol. 7, p. 108.

[12] Ordered at a meeting of all selectmen, October 2, 1677. *Ibid.*, p. 113.

[13] Vote at town meeting, March 11, 1700. *Ibid.*, p. 241.

[14] Vote at town meeting, June 12, 1706. *A Report of the Record Commissioners of the City of Boston Containing the Boston Records, 1700–1728* (Boston: Rockwell and Churchill, 1883), Vol. 8, p. 40.

[15] Vote in town meeting, May 9, 1711. *Ibid.*, p. 81.

[16] Vote in town meeting, August 27, 1711. *Ibid.*, p. 84.

[17] Walter Kendall Watkins, "The Great Street to Roxbury Gate, 1630–1830," *The Bostonian Society Publications*, Second Series, Vol. 3 (1919), p. 109.

[18] Vote of freeholders and other inhabitants of the town, March 1, 1717/8. *A Report of the Record Commissioners of the City of Boston Containing the Boston Records, 1700–1728*, Vol. 8, p. 40.

[19] Diary entry, August 30, 1717. "Diary of Samuel Sewell," *Collections of Massachusetts Historical Society*, Vol. VII, Fifth Series, p. 137.

[20] Meeting of freeholders, Sept. 22, 1746. City of Boston, *A Report of the Record Commissioners, Boston Town Records, 1742–1757* (Boston: Rockwell and Churchill, 1885), Vol. 14, p. 103.

[21] Watkins, *op. cit.*, p. 114.

[22] Gov. West to Lord Ashley, Albemarle Point in Ashley River, March 2, 1670/1. *Collections of the Historical Society of South Carolina*, Vol. 5 (1897), p. 267.

[23] Lord Ashley to Gov. West, quoted by Mrs. St. Julien Ravenel, *Charleston, the Place and the People* (New York: Macmillan, 1906), p. 13.

[24] W. Roy Smith, *South Carolina as a Royal Province, 1719–76* (New York: Macmillan, 1903), p. 196.

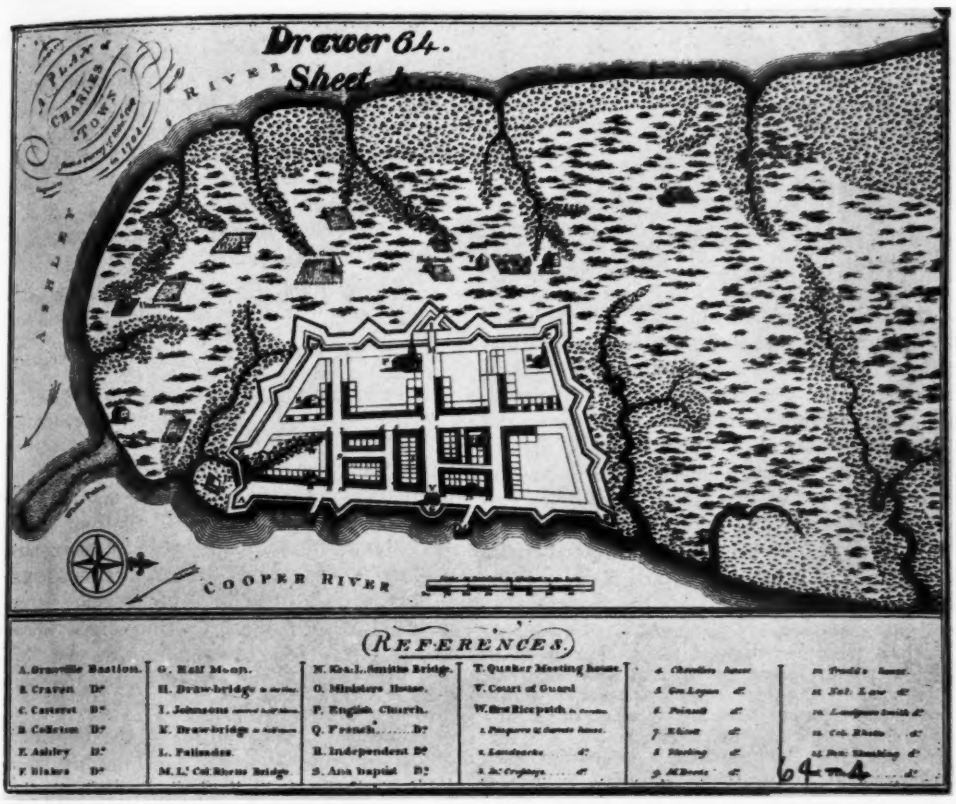


FIG. 2. Charleston. The Edward Crisp survey of Charleston. Although the date 1704 appears on the map, most authorities feel the state of the fortifications indicates a later date, possibly 1711. (From the National Archives, Record Group No. 77.)

work on "forts, half moons, platforms, entrenchments, flankers and parapets, sally ports, a gate and a drawbridge and blind."[25]

The result of this cumulative work is well illustrated by the Crisp Survey of 1704, shown here as Figure 2. Although the settlement was located on a peninsula, this site was rather large for the scale of the settlement, and a wall was built around the city. The fortifications extended along the present Water Street on the south, Meeting Street on the west, and generally along Cumberland Street on the north. A description of about the same period (1708) states that the city "is surrounded with a regular fortification, consisting of bastions, flankers and half moons, ditched and palisaded, and mounted with 83 guns."[26] A gate and a drawbridge were located in "Johnson's Half Moon" on the west, and it is thought the entrenchments were dug deep enough so that water stood in them.[27]

Charleston, as other cities, experienced difficulty in keeping its citizens away from the fortifications. An act of 1704 read in part:

> . . . some inconsiderate or evil disposed persons, not regarding or minding the evil consequences of defacing and breaking down the said fortifications, do presume to climb and get over the said intrench-

[25] Thomas Cooper and David J. McCord, eds., *Statutes at Large of South Carolina, 1662–1838* (10 vols.; Columbia: A. S. Johnston, 1836–41), Vol. VII, pp. 28–30.

[26] Grand Council to Board of Trade, Charleston, 1708. MS. letter quoted by William Rivers, *A Sketch of the History of South Carolina* (Charleston: MacArthur and Co., 1856), p. 207.

[27] *Yearbook, 1883, City of Charleston South Carolina* (Charleston: News and Courier Book Press, 1883).

ments and other works, and so break them down, and lay open the said town. . . .

. . . if any white person . . . above the age of 16, shall presume or endevour to climb or pass over any part of the said fortifications, inward or outward of Charlestown or go down into the ditch or trench, he . . . shall fine and forfet the sum of 20 shillings . . . or . . . shall recieve, at the public stocks, fifteen lashes on the bare back, and fifteen lashes more at the inward bridge. . . .[28]

Even so, repairs were ordered many times: in 1707, 1714, 1719, 1721, and 1725, for example.

Although the town gradually spread outside of the wall, particularly after 1717, only much later were efforts made to enlarge the fortifications. In 1745 a moat 36 feet wide and "as deep as possible," with an earth rampart, was ordered constructed completely across the neck at the north edge of the city.[29] It was destroyed by the hurricane of 1752.[30] Apparently in spite of several later plans and beginnings of construction, no later wall was ever completed. A long-time resident describes the situation as of 1763:

> The old Fortification, on the Land Side, is in Ruins; a new Work was begun in 1757, a little without the other; the plan was a Horn-Work, to be built with Tappy, and flanked with little Batteries and Redoubts, at proper Distances; the Whole to extend from one River to the other, but a Stop was put to this likewise, after considerable Progress was made in it, either for Want of Money, which is probable, or from an Opinion, that it was unnecessary.[31]

The remains of this "new work" were still apparent as late as 1774.[32]

SAVANNAH

Savannah, founded by Oglethorpe on the forty-foot Yamacraw Bluff overlooking the Savannah River in 1733, was considered to be part of a buffer colony against Spain. A plan for a garrison town, or perhaps more accurately a sort of "town of refuge," was drawn before the settlers left England. Francis Moore, storekeeper at Frederica, and knowledgeable companion of Oglethorpe, describes the layout of Savannah after a visit in 1736, and explains the origin of the distinctive squares of the city as follows:

> The use of this is, in case a war should happen, the villages without may have places in the town, to bring their cattle and families into for refuge, and to that purpose there is left a square in every ward, big enough for the outwards to encamp in. There is ground also kept round about the town ungranted. . . . for the fortifications whenever the occasion shall require.[33]

A stockade was started around the city immediately and a portion of it shows on the Peter Gordon map of 1734, made one year after the city was founded,[34] but it apparently was never completed, as it is not mentioned by Moore. He describes only a "guard house inclosed with palisades a foot thick where there are nineteen or twenty cannons mounted, and a continual guard kept by the freeholders."[35]

When war broke out with the Creeks in 1757, Governor Ellis felt the time had come to fortify Savannah and called in a professional to design the works, John Gerar William De Brahm, "His Majesty's Surveyor General for the Southern district of North America." De Brahm himself, whose plan appears as Figure 3, has described his proposal and accomplishments:

> The Author . . . proposed a well palisadoed Intrenchment to envelope the City, so as to make it a Receptacle and Shelter for all the Planters, their Families, Slaves, and . . . The Author's Advice met with general Approbation, and He laid two Poligons with three Bastions on the south side, and with four Poligons more (two on the east, and two others on the west side of the City, each ending with a demi-Bastion). He joined the Streams. The Soil of Savannah being a meer Sand, to make this

[28] An Act to prevent the Breaking Down and Defacing the Fortifications in Charles Town, Read three times and ratified in open Assembly, Nov. 4, 1704. Thomas Cooper and David J. McCord, *op. cit.*, p. 36.

[29] Act, ratified May 25, 1745. W. Roy Smith, *op. cit.*, p. 200.

[30] *Ibid.*, p. 201.

[31] George Milligen-Johnston, *A Short Description of the Province of South-Carolina* (London: John Hinton, MDCCLXX), p. 36, reprinted in Chapman J. Milling, ed., *Colonial South Carolina, Two Contemporary Descriptions* (Columbia: Univ. of South Carolina, 1951), 209 pp.

[32] "No gates have been hung nor guns mounted on it and it is now little more than a nuisance." "Charleston 1774 as Described by an English Traveler," *Historical Magazine and Notes and Queries Concerning the Antiquities, History and Biography of America*, Vol. IX, No. 11, Series 1 (Nov. 1865), p. 343.

[33] Francis Moore, entry dated Feb. 1736, in *Voyage to Georgia, Begun in the Year 1735* (London, 1744). Reprinted in *Collections of the Georgia Historical Society*, Vol. I (1840), p. 98.

[34] Peter Gordon map, 1734. Widely reproduced; see, for example, Charles Calcock Jones, *The History of Georgia* (Boston: Houghton, Mifflin and Co., 1883), Vol. I, p. 127.

[35] Francis Moore, *op. cit.*, p. 95.

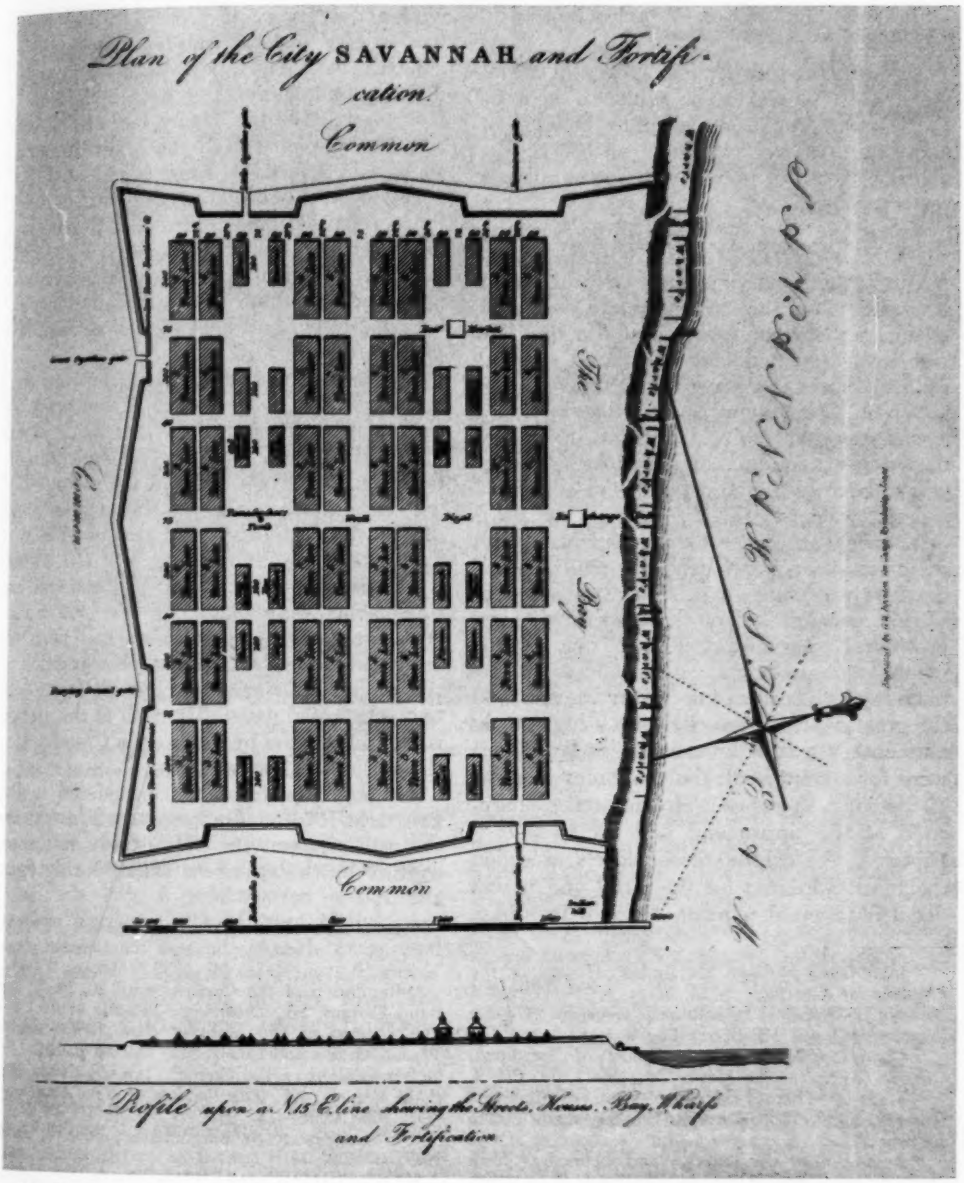


FIG. 3. Savannah. De Brahm, given the responsibility of fortifying Savannah in 1757, produced the above plan and supervised the construction of the wall. Map from John Gerar William De Brahm, *History of the Province of Georgia: With Maps of the Original Surveys* (Wormsloe [Plantation] Georgia: George Wymberley-Jones, 1849). (Courtesy of Savannah Public Library.)

keep in a breastwork, the Author was obliged to have the outside Talus faced with pine Saplings set in the Ground, and inclined their tops in from the Talus of the Scarp, the Governor improved this Intrenchment with adding wooden Tour's Bastionees to each corner Bastion, one of which was placed in the Angle of the Bastion to serve as Cavalier's couverte's, with strong Platforms in their

first Stories for Cannons of twelve Pounders to range and command the Country.[36]

Although the fortifications were still incomplete when Gov. Wright assumed the reins (1760), work progressed rapidly and Savannah soon provided an asylum against the Indians.[37] No trace of the wall seems to have remained by 1770.[38]

NEW YORK CITY

The Dutch, choosing a site for their New Amsterdam on the lower tip of Manhattan Island, perhaps consciously found a spot that was easily fortified. Situated amid growing English power, its need for defense became apparent. It is not surprising, therefore, that a wall was soon built, a wall that because of the later surroundings of its namesake street was to become the most famous of all American city walls.

Although an earlier barrier had been built some nine years previously as a cattle fence,[39] the first wall came as a result of reports of warlike preparations of the New Englanders in 1653. As the fort could not "hold all the inhabitants nor defend all the houses" in the town it was considered necessary "to surround the greater part of the city with a high stockade and a small breastwork."[40] Ambitions were for a stout wall; the "palisades must be 12 feet long, 12 inches in circumference, sharpened at the upper end and set in line."[41] However, the cost of these pickets exceeded the funds allocated for the job,[42] and it was decided to build a plank fence 9 feet high with a post every 15 feet.[43] In addition, it was ordered that the wall be strengthened by a moat: "Citizens should begin immediately digging a ditch five feet deep and 11 or 12 feet wide at the top."[44] By July of that year all was complete,[45] and the lower tip of the island was effectively barricaded.

However, only a year later Stuyvesant called for its repair.[46] The citizens seem to have had little respect for their defenses. It is reported that "certain persons having cut up and burned a number of the city's palisades" they are required to cut and haul 100 palisades in lieu of those they burned.[47]

In September of 1655, the "dangerous conditions of the times" indicated the need for further strengthening of the defenses. A "curtain of Planks" five to six feet high was ordered "nailed to the sides of the palisades,"[48] and the improvement was finished in eight days.[49] In the next year it was ordered that "a large and suitable gate be fixed in the works near the East River."[50] Continual improvements were made. By the autumn of 1660 the old line of palisades had been rebuilt and strengthened by the construction of six masonry bastions, and the city now had two substantial gates. The wall in this period is well illustrated by the famous Castello Plan shown here as Figure 4.

Although the city was surrendered to the British in 1664 no attempt was made to storm the wall, nor was the wall a factor in the recapture by the Dutch in 1673. During these

---

[36] John Gerar William De Brahm, *History of the Province of Georgia: With Maps of the Original Surveys* (Wormsloe [Plantation] Georgia: George Wymberley-Jones, 1849), pp. 38–39.

[37] Charles Calcock Jones, *History of Savannah, Georgia* (Syracuse: D. Mason Co., 1890), pp. 171–2.

[38] Map by Thomas Shouder, Dept. of Surveyor General, Feb. 1770. Copy in the Savannah Public Library.

[39] Resolution of the Director and Council of New Netherlands, March 31, 1744. *Colonial Manuscripts* (Albany: New York State Library, n.d.), Vol. IV, p. 186.

[40] Resolution of the Council, March 13, 1653. Berthold Fernow, ed., *The Records of New Amsterdam from 1653 to 1674* (New York: Knickerbocker Press, 1897), Vol. I, p. 66.

[41] Resolution of the Council, March 17, 1653. *Ibid.*, p. 72.

[42] Resolution of the Committee, March 18, 1653. *Ibid.*, p. 73.

[43] Resolution of the Committee, March 19, 1653. *Ibid.*, p. 73. Contract between the Committee and Thomas Barton, March 20, 1653. *Ibid.*, pp. 73–74.

[44] Resolution of the Council, April 20, 1653. Berthold Fernow, ed., *Documents Relating to the Colonial History of the State of New York* (Albany: Weed, Parsons and Co., 1883), Vol. 14, p. 201.

[45] Resolution of the Council, July 28, 1653. Berthold Fernow, ed., *The Records of New Amsterdam from 1653 to 1674*, p. 90.

[46] P. Stuyvesant to Burgomasters, June 8, 1654. *Ibid.*, p. 209.

[47] Calligraphic Historical MS., Dutch, 145, quoted in I. N. E. Stokes, *The Iconography of Manhattan Island* (New York: R. H. Dodd, 1915–28), Vol. 4, p. 155.

[48] Resolution of the Council, Sept. 20, 1655. Berthold Fernow, ed., *The Records of New Amsterdam from 1653 to 1674*, p. 363.

[49] Burgomasters to Councillors, Sept. 28, 1655. *Ibid.*, p. 365.

[50] Burgomasters to Council, Feb. 1, 1656. *Ibid.*, Vol. II, p. 29.

FIG. 4. New York. The Castello Plan of 1660 shows the fortifications of that date, with the wall on the land side at the line of the present Wall Street. (Print of redraft of the Castello Plan courtesy of the Phelps Stokes Collection, New York Public Library.)

later hostilities, with the Dutch in temporary possession, the wall was still an effective barrier and considered to be important in the defense of the city. A proclamation in that year by Gov. General Colve of New Netherlands reads as follows: "The City gate shall be closed at sundown by the Mayor of this city . . . and opened at sunrise," and no one save "the watch alone [shall] attempt coming from sunset to sunrise on the bulwarks, bastions or batteries [of the city] on pain of bodily correction," and it is strictly forbidden for anyone to come into the city or go out of it "in any other manner, way or means than through the ordinary City gate on pain of DEATH."[51] But when the British took over the next year, again the wall was not a factor.

Although the wall was altered and repaired several times in the next thirty-five years, by 1699 it was in poor condition and was ordered removed. The "former line of fortifications that did range along Wall Street from the East to the North River, together with the bastions, are fallen to decay . . . the fortifications may be demolished . . . and the stones of ye bastions . . . may be appropriated to the building of said City Hall."[52]

Thus ended New York's famous wall. But the concept was to be called up again in times of danger. In 1706 work was begun on another wall. "We have above 200 at work every day on the Fortifications. . . . The Northside of this city from Hudson's River to the East River will be inclosed with Stockadoes and a very good Breastwork. . . ."[53] The lo-

[51] Order of the Governor General to the Burghery, Dec. 27, 1673. *Ibid.*, Vol. VII, pp. 35–36.

[52] Petition of the Council, August 18, 1699. *Minutes of the Common Council of the City of New York, 1675–1776* (New York: Dodd, Mead and Co., 1905), Vol. II, p. 82.

[53] News from New York, in *Boston News-Letter*, May 27—June 3, 1706, quoted in I. N. E. Stokes, *op. cit.*, pp. 455–56.

FIG. 5. New York. David Grim, in 1813, produced a map depicting the city as it was in 1742–44. In spite of the time lag the map has a high reputation for accuracy. The line of the wall across the peninsula is generally that of the present Chambers Street. (Courtesy of the New York Historical Society, New York City.)

cation and duration of this wall is uncertain.[54] Later, in 1745, a line of palisades was erected across the island, generally at the present Chambers Street. Grim remembers it made "of cedar logs 14 feet long, 9 or 10 inches in diameter . . . placed in a 3 foot trench. . . . In the line of Palisades were 3 Blockhouses about 30 feet square . . . there were 4 large gates at Pearl, Chatham, Broadway and Greenwich Streets."[55] His map of this wall, drawn from memory many years later, is shown as Figure 5. The greater part of it was reported still standing in 1756, "as a monument of our folly, which cost the province about 8,000 pounds."[56]

[54] D. T. Valentine, *Manual of the Corporation of the City of New York, 1862* (New York, 1862), p. 523.

[55] Statement of David Grim accompanying his map drawn in 1813, depicting city as of earlier period. I. N. E. Stokes, *op. cit.*, Vol. I, p. 272.

[56] William Smith, *The History of the Province of New York* (2nd ed.; Philadelphia: Mathew Carey, 1792), p. 206. Comment written June 15, 1756.

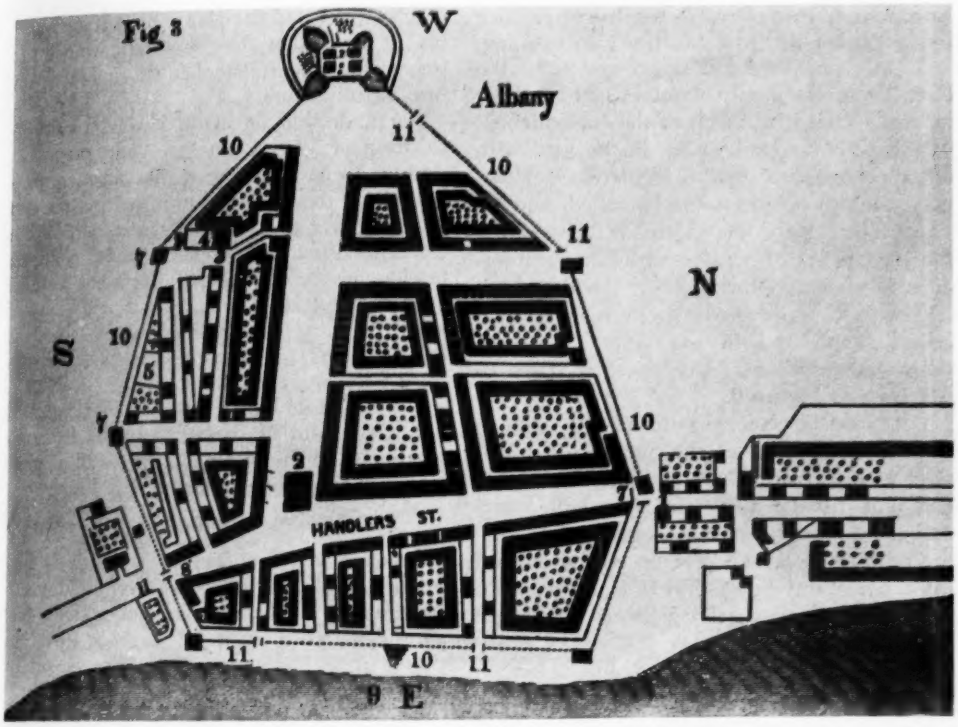


FIG. 6. Albany. Rev. John Miller visited Albany in 1695 and produced the above map. Map reproduced from John Miller, *A Description of the Province and the City of New York*, republished in *Gowans' Bibliotheca Americana* (New York: Gowans, 1845), p. 265.

ALBANY

Albany was a walled city for at least a hundred years. The "church neighborhood" in the vicinity of Fort Orange that was to become Albany was first ordered enclosed in November 1659, as a "much needed" defense "against attacts by the savages in these dangerous times." The wall was to be "made of posts and boards, to wit, eight boards high, with seven bastions to protect the curtins, which [fence] shall surround the greater part of the village . . . the length of its circumference being 250 rods."[57] At the end of the month the defense was ordered continued on the river side with a "fence of posts and boards 7 to 8 feet high. . . ."[58] By July the fence was complete and Stuyvesant authorized a tax of three guilders on every chimney to pay for it.[59]

Although palisades of wood needed constant replacement, visitors to Albany in the following decades regularly characterize it as a walled city. In 1680 travelers from Europe describe Albany as a town "with several good streets, on which there may be about eighty or ninety houses. . . . The town is surrounded by palisades and has several gates corresponding with the streets."[60] To the French, with good intelligence reports, the town appeared (1689) "as large as Montreal, surrounded by picquets, at one end of which is an earthan Fort: said town may contain 150 houses. . . .

[57] Court order, Fort Orange, Nov. 4, 1659. A. J. F. Van Laer, trans. and ed., *Minutes of the Court of Fort Orange and Beverwyck, 1657–1660* (Albany: Univ. of the State of New York, 1923), Vol. II, p. 226.

[58] Court order, Fort Orange, Nov. 24, 1659. *Ibid.*, p. 229.

[59] Order of P. Stuyvesant, Fort Orange, July 25, 1660. *Ibid.*, p. 287.

[60] Journal entry, May 30, 1680. Bartlett B. James and J. Franklin Jameson, eds., *Journal of Jasper Danckaerts, 1679–1680* (New York: Scribners, 1913), p. 17.

I propose, in case of resistance, to cut or burn the palisades, in order to afford an opening, and [to] enter sword in hand and seize the fort. These being only about 14 feet high can be easily scaled by means of the conveniences we can find, or by blowing in the gate with a few petards or 2 small field pieces."[61] However, the attack was never launched, and six years later (1695) Rev. John Miller visiting the city speaks of a place with "a circumference of about six furlongs . . . [with] about 200 houses. . . . The whole city is well stockadoed round and in the several fortifications . . . are about thirty guns."[62] His map is shown as **Figure 6.**

But constant repair was necessary much to the discouragement of the authorities downriver in New York. In 1701 "The breastwork and pallisadoes which were round the town and the two stone bastions on the land side are quite demolished. They were never well built, nor of any great service."[63] Three years later, however, new palisades were set around the whole town.[64]

Fifty years later, the mayor and aldermen were petitioning the authorities in New York:

> Whereas the city of Albany is a frontier town . . . at present altogether undefensible, open and exposed . . . your petitioners . . . would be pleased to pay a tax of 6,000 pounds on Estates . . . to defray the expense of building a wall with Bastions or Batteries . . . . There was a stone wall begun several years ago, it was carried on but a little way; from the wall Stockadoes were put up and Block houses erected at convenient distances around the city. Many of the stockadoes are now rotten and falling down . . . [The stone wall] could be finished.[65]

In April of 1754 three men went to New York City and pressed for action on a stone wall,[66] but in September they reported "we had . . . begun to procure stockadoes and are very busy setting same. . . ."[67]

For a decade or so this wall served well. In June of 1755 Cornelius Van Dusen "undertook to open and shut the gates . . . for a space of 2 months."[68] Smith, writing the next year, reports that "the greatest part of the city is fortified only by palisades."[69] A French intelligence report states: "It is not fortified on the forest side except by an enclosure of walls or pickets, without a ditch, which is flanked at certain distances; the river defends the enclosure at the other side."[70]

Twenty years later (1775) all that remained was a segment of the stone wall: "from the north gate to the water is a fine wall of stone and lime with loop holes."[71]

SCHENECTADY

Schenectady, founded as a village of four square blocks in 1662, was reportedly surrounded by earth embankments and ten-foot palisades with two gates as early as 1670.[72] A visitor in 1680 describes "a square set off by palisades. There may be about 30 houses. . . ."[73]

That the wall was necessary, but no substitute for vigilance, was demonstrated when the French and Indians burned the city in February, 1690. The French, reporting on the incident, wrote:

> The Town . . . forms a sort of oblong with only two gates—one opposite the road we had taken, the other leading to Orange. Messieurs de Sainte

---

[61] Chevalier de Callieres to the Marquis of Svignday, Montreal, January, 1689. E. B. O'Callaghan, ed., *The Documentary History of the State of New York* (Albany: Weed, Parsons and Co., 1850), Vol. I, pp. 180–81.

[62] Rev. John Miller, *A Description of the Province and City of New York as They Existed in the Year 1695* (London: Thomas Rodd, 1843), p. 6.

[63] Livingston to the Lords of Trade, New York, May 13, 1701. E. B. O'Callaghan, ed., *Documents Relative to the Colonial History of New York* (Albany: Weed, Parsons and Co., 1854), Vol. IV, p. 878.

[64] Congreve to the Lords of Trade, New York, July, 1704. *Ibid.*, Vol. IV, p. 1128.

[65] Petition of mayor and aldermen, Albany, May 29, 1753. J. Munsell, ed., *Collections on the History of Albany* (Albany: Munsell, 1865), Vol. I, p. 81.

[66] Resolution of the Common Council, April 4, 1754. *Ibid.*, p. 88.

[67] Letter from Common Council to Lt. Governor, Sept. 11, 1754. *Ibid.*, p. 91.

[68] Resolution of the Common Council, June 11, 1755. *Ibid.*, p. 96.

[69] June 15, 1756. William Smith, *History of the Province of New York* (2nd ed.; Philadelphia: Mathew Carey, 1792), p. 216.

[70] "Topography of Country between Oswego and Albany, 1757" (unsigned), E. B. O'Callaghan, ed., *op. cit.*, Vol. X, p. 677.

[71] Diary entry, Albany, Oct. 6, 1775. Diary of Rev. Samuel Chandler quoted in Munsell, *op. cit.*, Vol. II, 1867, p. 373.

[72] Edwin G. Conde, Chap. III, in Myron F. Westover, *Schenectady Past and Present* (Strasburg, Va.; Shenandoah Pub. House, 1931), p. 12.

[73] Journal entry, April 26, 1680. Bartlett B. James and J. Franklin Jameson, eds., *Journal of Jasper Danckaerts, 1679–1680* (New York: Scribners, 1913), p. 213.

A PLAN
of
SCHENECTADY.
Situated Lat. 43. Long. 74. 30.
A. Schenectady
B. Wooden fort with four Block houses for flankers
C. Block houses to defend the Stockades
D. Stockades planted round the Town
E. The Nearest high Ground to y Town which is about two Y. from it
F. Part of a line of an Entrenchment thrown up and of seven Battery to shew how such works are Constructed
G. Barracks or Sheds where part of y Reg. were Lodged last Winter
The Boundary on each side of the River is pretty nearly on a Level
Except where it's marked otherwise on the Draught

FIG. 7. Schenectady. This map of Schenectady is attributed to British Army topographers and is thought to date from about 1750. Map found in Jean Rocque, *A Set of Plans and Forts in America Reduced from Actual Surveys* (London, 1763), p. 8. (Courtesy of the New York State Library, Albany.)

> Helene and de Montel were to enter the first which the squaws pointed out, and which in fact was found wide open . . . there were upwards of 80 well-built houses in the town.[74]

However, the town was rebuilt on about the same plan. When Rev. John Miller visited it in 1695 it was "quadrangular, with a treble stockado, a new blockhouse at every angle."[75] His map of the settlement shows a wall. Here, too, constant repair was necessary as "the timber in this country is rotten, when exposed to the weather, in six or eight years time."[76] In 1702 a report was made that "Schenectady is now an open village. . . . It was formerly stockadoed round but since the peace, no care having been taken to repair the Stockadoes, they are all down."[77] But in 1704 it was enlarged by moving out the stockade on one side. "The stockadoes set up on the West side of the town of Schenectady [are] to be removed from the place where they now stand and put up as near the river as the ground will [allow]. . . ."[78]

Half a century later Schenectady still had its stockade, as a map made about 1754 shows (Fig. 7). A French intelligence report of 1757 describes the city fairly accurately:

> Chenectedi, situated on the bank of the Moak river, is a villiage of about 300 houses. It is sur-

[74] An account of Mons. de Monseignot, Comptroller General of the Marine in Canada. E. B. O'Callaghan, ed., *The Documentary History of the State of New York* (Albany: Weed, Parsons and Co., 1850), Vol. I, p. 187.

[75] Rev. John Miller, *op. cit.*

[76] Livingston to the Lords of Trade, New York, May 13, 1701. *Op. cit.*

[77] Lord Cornbury to the Lords of Trade. Chear Hall, Orange County, New York, September 24, 1702. *Ibid.*, p. 968.

[78] Order of Ed. Viscount Cornbury, Gov. of New York, July 29, 1704. *New York Colonial Manuscripts* (Albany: New York State Library), Vol. 49 (1704), p. 149a.

FIG. 8. Detroit. A view of Detroit from the river July 25, 1794. This is a wash drawing by an unknown artist, E. A., and was a gift of Lady Nancy Astor to the Burton Historical Collection, Detroit Public Library. (Courtesy of the Detroit Public Library.)

> rounded by upright pickets, flanked from distance to distance. Entering the villiage, by the gate on the Fort Hunter side, there is a fort to the right which forms a species of citadal in the interior of the villiage itself.[79]

Defenses of the village, it is reported, decayed after 1763. After the Revolution the stockadoes rotted and fell to the ground and Schenectady became an open village.[80]

DETROIT

Detroit was founded by the French as Fort Poncentrain in 1701, and, although early records are scanty, it seems to have had a wall for most of the first 120 years of existence. Cadillac brought with him about one hundred Frenchmen, half of them soldiers and half artisans, and "built a fort with four bastions of good oak pickets, fifteen feet long, stuck three feet in the ground."[81] The Detroit historian, Burton, describes enlargements in 1703 and particularly in 1707, although the village was only about 720 feet by 250 feet at that time.[82]

Maps by De Lery in 1749 and 1754, as well as a map in *Bellin's Atlas*, published in 1760, show Detroit remaining a small walled village. It had four east-west streets, St. Joseph, St. James, St. Ann, and St. Louis, and appears to have had bastions at each corner and several gates.[83]

The English took possession in 1760, and in a letter written the following year the place is described as follows:

> The palisades are in good repair. There is a scaffolding round the whole, which is only floored towards the land for want of plank, it is by way of a banquette. There are seventy or eighty houses in the fort, laid out in regular streets. . . .[84]

Detroit's wall was reportedly in a dilapidated condition when the town was surrendered to the Americans in 1796.[85] Views of the wall at this period are shown in Figures 8 and 9. Eight years later, 1804, when it was reported that part of the pickets and gates had been removed, the freeholders voted 13 to 12 against replacing them.[86] The next year the village was destroyed by fire, the entire picket

[79] Topography of Country between Oswego and Albany, 1757. O'Callaghan, ed., *op. cit.*, Vol. 10, 1858, p. 677.

[80] Howell Munsell, *History of the Country of Schenectady, New York, from 1662–1886* (New York: W. W. Munsell, 1886), pp. 35, 38.

[81] Letter of Chevalier de Callieres, Oct. 4, 1701. Silas Farmer, *The History of Detroit and Michigan* (Detroit: Silas Farmer and Co., 1889), p. 221.

[82] M. M. Quaife, ed., *When Detroit Was Young, Historical Studies by Clarence Monroe Burton* (Detroit: Burton Abstract and Title Co., 1951), p. 18.

[83] De Lery Maps in Burton Historical Collection, Detroit Public Library. For reproductions see *ibid.*, pp. 21 and 23. Map from *Bellin's Atlas* reproduced in the same volume, p. 25.

[84] Letter of Donald Campbell, 1761. Quaife, *op. cit.*, p. 144.

[85] Quaife, *op. cit.*, p. 28.

[86] Minutes of Town Meeting, April 28, 1804. *Corporation of the Town of Detroit, Act of Incorporation and Journal of the Board of Trustees, 1802–1815*, Burton Historical Collection, Detroit Public Library (Detroit, 1922), p. 56.

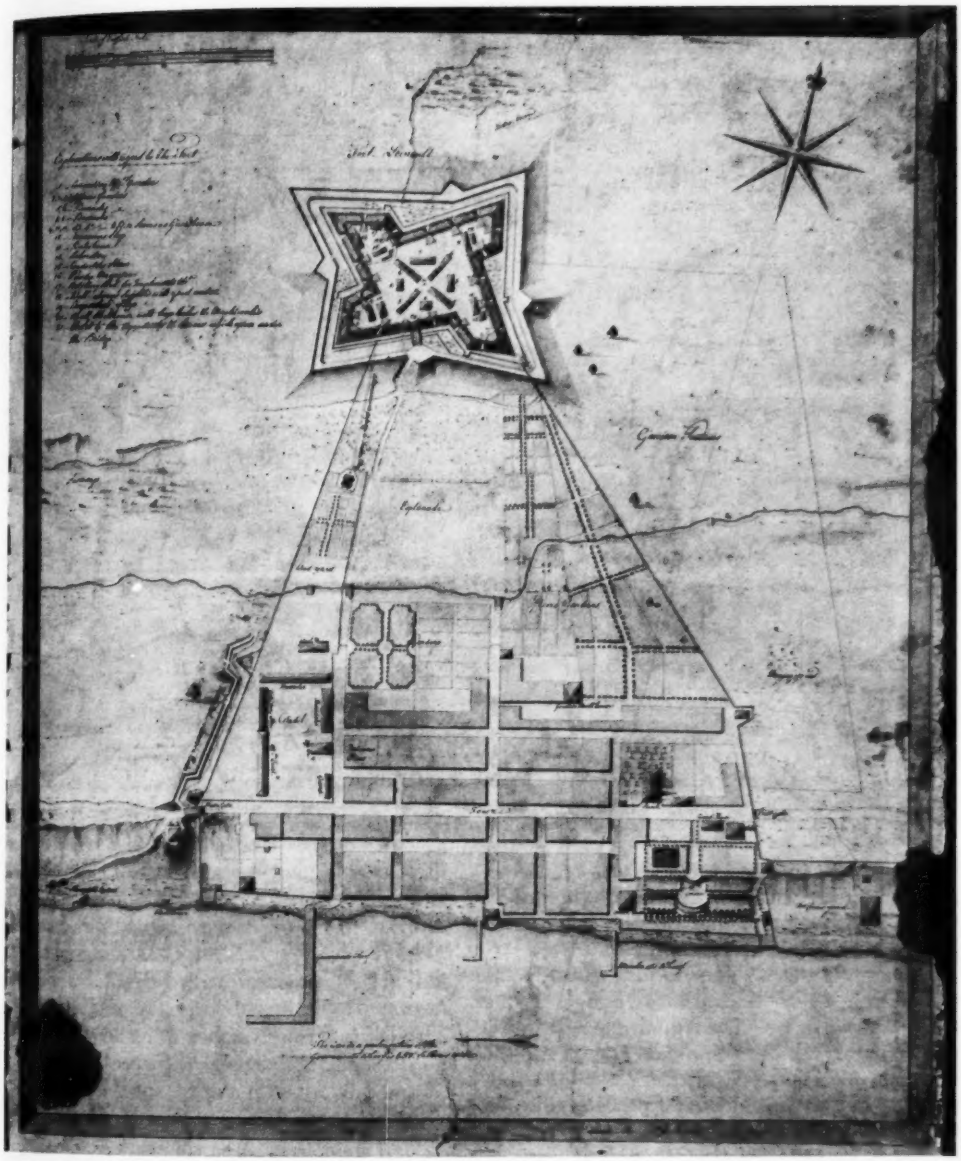


FIG. 9. Detroit. A recently discovered map of Detroit, drawn by Major J. J. U. Rivardi, of the Corps of Engineers, U. S. Army, dated March 29, 1799. (Courtesy of the Clements Library, University of Michigan.)

line was reported in ruins, and in rebuilding it was possible to enlarge the boundaries of the city.[87]

In 1807, however, a new stockade was begun to enclose the town on the three land sides.[88] A hundred militia men were put to

[87] Quaife, *op. cit.*, p. 30.

[88] Hull to Dearborn, Detroit, August 4, 1807. *Michigan Pioneer and Historical Collections*, Vol. 40 (1929), p. 170.

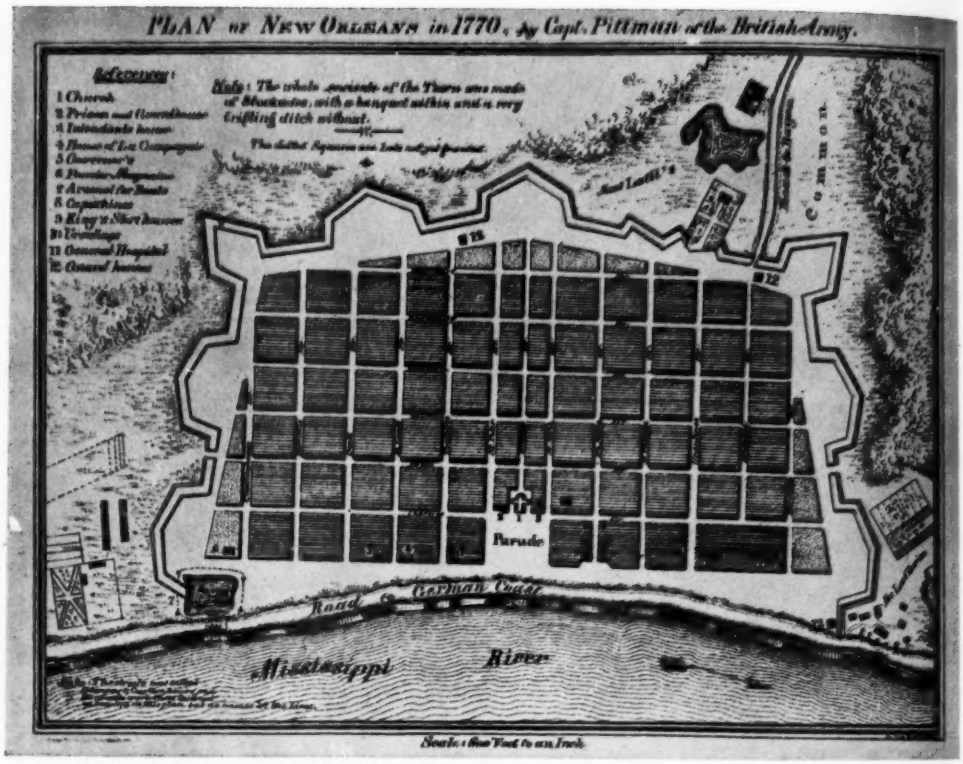

PLAN OF NEW ORLEANS in 1770, by Capt. Pittman of the British Army.
References:
1 Church
2 Prison and Guardhouse
3 Intendants house
4 House of the Company
5 Governor's
6 Powder Magazine
7 Arsenal for Boats
8 Capuchines
9 King's Store house
10 Hospital
11 General Hospital
12 Grand house
Note: The whole curtain of the Town was made of Blockwork, with a banquet within and a very trifling ditch without.
The dotted Squares are Lots not yet granted
Road
Parade
German Coast
Mississippi River
Scale of One Foot to an Inch

FIG. 10. New Orleans. This plan of New Orleans as of 1770 was drawn by Capt. Pittman of the British Army. Philip Pittman, *The Present State of the European Settlements on the Mississippi* (London: J. Nourse, 1770), map opposite p. 10. (Courtesy of the Chicago Historical Society.)

work on it.[89] By December it was complete, and although Governor Hull thought it was too large, he reports that no one wanted his house to be left outside of the wall.[90]

Apparently the wall lasted until about 1826. On March 12 of that year a notice in the *Gazette* proclaims: "The Adjourned sale of the pickets near the fort will take place on the Ground on Thursday next. . . ."[91] Farmer reports that at the sale 6,000 pickets, forming the fort and part of the stockade, were sold for from $2.00 to $3.00 per hundred.[92]

NEW ORLEANS

New Orleans, laid out by Bienville as a large rectangle in 1718, is shown on early plats as designed for a wall from the first.[93] Although unfortified for a decade,[94] it eventually was to have a system of walls and moats perhaps unsurpassed among American cities, and was to remain a walled city for about 75 years.

Following the Natchez Massacre, November 28, 1729, a palisade reportedly was built around the city with a moat, the whole flanked by blockhouses.[95] Thirty years later, after the

[89] Hull to Dearborn, Detroit, August 13, 1807. *Ibid.*, p. 179.

[90] Hull to Dearborn, Detroit, December 29, 1807. *Ibid.*, p. 243.

[91] *Gazette* (Detroit), March 13, 1827.

[92] Silas Farmer, *op. cit.*, p. 225.

[93] "Plan de la Ville de la Nouvelle Orleans (Janiver, 1723)." Original in the Ayer Collection, Newberry Library, Chicago. Widely reproduced: for example, see Stanley C. Arthur, *Old New Orleans* (New Orleans: Harmanson, 1936), facing p. 18.

[94] "Plan de la Nouvelle Orleans Capitale de la Louisiana 1728." Original in the Library of the Chicago Historical Society. Widely reproduced; see *ibid.*, p. 23.

[95] H. Mortimer Favort, "Colonial Forts of Louisiana," *Louisiana Historical Quarterly*, Vol. XXVI (July, 1943), p. 730.

fall of Fort Duquane (1760), it was considered prudent to surround New Orleans again with a ditch and a palisade.[96] The palisade was completed before the end of the year, but needed artillery for effective defense.[97]

Palisades, as usual, rotted quickly. After the Spaniards took control of the territory it was reported (1767): "The fortification of New Orleans consists of a stockade which is continually deteriorating on account of the moisture of the soil."[98] Again, in 1769, a report concerning the city "whose moat and palisade, constructed in the year 1760, cost the King of France 69,502 pesos *fuertes*," says the palisade is rotting.[99] Yet a plan of this period was published with the comment, "The whole enciente of the Town was made of Stockadoes, with a banquet within and a very trifling ditch without."[100] The plan, by Capt. Pittman of the British Army, is reproduced as Figure 10.

Repairs seem to have been made constantly, however, for 25 years later, in 1794, Carondelet, the Spanish Governor, discussed the fortifications of the city:

> New Orleans, whose defences consisting only of five earth redoubts, with a trench and a covered road furnished with a strong stockade, these redoubts being united with one another by the same stockade protected by a good glacis, at whose foot is a trench and a ante-glacis, with a redan in the middle of each curtin. . . .[101]

If provided with troops he thought it could be held.

A few years later, the city was visited by Victor Collot, a brilliant military geographer on a trip down the Mississippi, apparently serving as a French spy. He was held in jail in New Orleans but was eventually deported with his maps intact, and was able to describe the city:

> The form of the town is that of an oblong square, five streets in breadth and seven in length, each at right angles. The population is composed of about ten thousand souls, including the free mulattoes and slaves.
>
> The five forts and the battery . . . are connected by a ditch forty feet broad and seven deep. With the earth of the ditch had been formed on the inside, a causeway three feet in height, and on which were placed great picquets of twelve feet, very near each other. Behind these picquets is a small banquette. On the side of the ditch the earth has been simply thrown out, which renders the slope gentle and easy. There is always four feet of water [in the ditch] even in the driest seasons.[102]

However, a decade later, in the spring of 1803, before the Americans arrived, the barrier had greatly deteriorated:

> The fortifications have never been kept up, and are falling into decay; the ditches are filling up; the terraces are crumbling down, the palisades are wanting, or rotten; the bridges have given way, or consist of only one or two beams; the gates are off their hinges, and are lying on the ground.[103]

SAINT LOUIS

St. Louis, the exposed headquarters of the Spanish Colonial Administration of Upper Louisiana from 1770 to 1804, was walled for a brief period during the American Revolution. The only attack came, however, before the wall was built.

The settlement heard rumors of an impending attack by Indian allies of the British in the spring of 1780. The governor ordered four stone towers or forts constructed, but with time and money running out he settled for a trench dug around the town connecting the one completed tower to the river on both ends of the settlement. The finished trench was a little over a mile long. Although crude, the defenses held during the attack (May 26, 1780). Persons caught in the fields, outside the line of the trench, were massacred.[104]

---

[96] Resolution of Court Martial, New Orleans, Jan. 2, 1760. Charles Gayarre, *History of Louisiana, The French Domination* (New York: Redfield, 1854), Vol. II, p. 88.

[97] Kerlerec to superiors, Dec. 21, 1760. *Ibid.*

[98] Governmental Expenses, Colony of Louisiana, 1767. Lawrence Kinnaird, ed., "Spain in the Mississippi Valley 1765–1794." *Annual Report of the American Historical Association, 1945*, Vol. II, Part 1, p. 17.

[99] O'Reilly to Arriaga, New Orleans, Dec. 29, 1769. *Ibid.*, p. 147.

[100] *London Magazine*, Vol. 30 (April, 1761), p. 209.

[101] Carondelet to Alcudia, New Orleans, Nov. 1794. Frederick J. Turner, ed., "Carondelet on the Defense of Louisiana, 1794," *American Historical Review*, Vol. II, No. 3 (April, 1897), p. 484.

[102] Victor Collot, *A Journey in North America* (Paris: Arthur Bertrand, 1826), Vol. 2, pp. 95–96.

[103] Colonial Prefect Laussat to Minister Decres, New Orleans, 1803. Charles Gayarre, *History of Louisiana, the Spanish Domination* (2nd ed.; New Orleans: Gresham, 1879), Vol. III, p. 582.

[104] De Leyba to Galvez, St. Louis, June 8, 1780. A. P. Nasatir, "St. Louis During the British Attack of 1780," *New Spain and the Anglo-American West, Historical Contributions Presented to Herbert Eugene Bolton* (Los Angeles: Privately published, 1932), Vol. 1, pp. 245–6.

Scale ½ of the original map
ST. LOUIS IN 1780
from the original map made by
AUGUSTE CHOUTEAU,
and now on file in the U.S. Recorders office
constructed and engraved for
CAMPBELL'S GAZETTEER OF MISSOURI.
EXPLANATION,
The shaded portions represent holes in the original map.
Aug. Chouteau.
Saint Louis des Illinois

FIG. 11. St. Louis. Auguste Chouteau designed the fortifications for St. Louis in 1780. This redraft of his original map was published five years later. Robert Allen Campbell, ed., *Campbell's Gazetteer of Missouri* (St. Louis: R. A. Campbell, 1874). (Reproduced from the collections of the Library of Congress.)

Danger from the same source persisted and a substantial wall was planned. The new governor, Cruzat, describes the plan to his superior:

> I therefore feel obliged . . . to fortify the town all around with a stockade of ten feet in height, and six inches thick. The only part not to have it will be that fronting the river, for here it is fortified by nature. This royal work I have already begun, and expect to have completed before the coming of our enemies . . . .[105]

Chouteau's map is reproduced as Figure 11.

Work on the wall reportedly went forward during the winter and was completed in the summer of 1781:

> Bastions were built at the two western angles to cover the curtins. This great stockade, over a mile in length and with four principal gates was completed by the summer of 1781—over a year after the attack, but still two years before the signing of the peace.[106]

But the stockade rotted away in less than four years,[107] much to the disgust of Lieut. Gov. Perez, who proposed a stone wall.[108] Although given permission to go ahead, nothing further was done. His comments may be considered typical of the European attitude toward the wooden pickets used to stockade American cities:

> It is not surprising that the stockade which was constructed in the time of my predecessor had become useless in so short a time, for it is a defect in the quality of the timber of this country which decays very readily within four or five years . . . . Absolutely [nothing] has remained . . . except the tower of San Carlos and the demilune . . . .[109]

No further walls were built, although many plans were made, and a ring of stone towers was eventually constructed.[110]

[105] Cruzat to Galvez, St. Louis, Dec. 18, 1780. Lawrence Kinnaird, ed., "Spain in the Mississippi Valley, 1765–1794," *Annual Report of the American Historical Association, 1945*, Vol. II, Part I, p. 410.

[106] "Hunt's Minutes," July 29, 1825. Typed copy in library of Missouri Historical Society, pp. 81–82.

[107] "Cruzat's Plan of St. Louis, 1780." Marginal note by Perez on map, reproduced and translated in Louis Houck, *The Spanish Regime in Missouri* (Chicago: Donnalley and Sons, 1909), Vol. I, pp. 273–4.

[108] Perez to Miro, St. Louis, Dec. 2, 1788. *Ibid.*, p. 271.

[109] *Ibid.*

[110] Charles E. Peterson, "Colonial St. Louis, Part III," *Bulletin of the Missouri Historical Society*, Vol. 4 (October, 1947), pp. 29–30.

SALT LAKE CITY

The Mormon cities of Utah, although founded at a relatively late date, were isolated from the rest of the country, and, after Indian scares in 1853, the settlements were ordered to "fort in." Many of the settlements completed substantial walls.

In Salt Lake, it was decided to build a wall around three sides of the city leaving the Jordon River on the west side as a natural barrier. Even so, the wall was to have a length of 2,731 rods[111] (about 8.5 miles), by far the longest and most ambitious undertaking of any of America's walled cities. Not only were the walls to be long, they were also massively designed. The *Deseret News*, Oct. 1, 1853, published the committee's recommendation to the council:

> We recommend . . . that the wall be made 12 feet high and 6 feet thick at the bottom, carried up with an equal slope on each face so as to be 2½ feet thick at the height of six feet, then carried up the remaining six feet 2½ feet thick and rounded at the top; that it be made of the earth taken from the ditch, which should be properly moistened mixed with weeds, sagebrush, small willows, grass, cane, bullrushes . . . and gravel where convenient, and placed on the ground for the wall at a depth suited to its consistency, and when the first layer is dry, repeat the operation until the wall is completed; that the ditch be dug large enough to furnish the earth for the wall.
>
> Of necessity there must be gates and bastions . . . five or more gates are recommended . . . that half-moon bastions be made into the wall at distances as near 40 rods apart as the nature of the ground will permit, and at the gates and corners.[112]

This proposal was submitted to the council and was adopted with a tax of $10.00 a lot to finance the wall.[113]

During the next year the *News* printed many reports of progress, exhortations to work on the wall, statements of progress in other cities, and accounts of ward parties working on the fortifications. The following item catches the flavor of the times: "The 18th Ward is progressing rapidly with the city wall, on the eminence N.E. of the Temple lot. What are the other wards [doing] . . . about this pleasent weather? Let spades answer."[114]

[111] Report of Committee to Mayor and City Council, Aug. 27, 1853. *Deseret News*, Oct. 1, 1853, p. 2.
[112] *Ibid.*
[113] Resolution of the City Council, Aug. 29, 1853. *Ibid.*
[114] *Deseret News*, Nov. 24, 1853.

Although officially reported as "fast progressing" and "it is expected it will be finished the present season," there apparently is no record of total completion.[115] However, travelers coming through the city in 1855 describe "a mud wall . . . almost entirely encircles this rising city, to which it serves as a rampart."[116] In a few years (1860), according to the best eye-witness account available, the wall seems to be deteriorating:

> On our way we passed the vestiges of an old moat from which was taken the earth for the bulworks of Zion—a Romulian Wall, of puddle, mud, clay and pebbles . . . . A glance will show that this much talked of fortification is utterly harmless; it is commanded in half a dozen places; it could not keep out half a dozen sappers for a quarter of an hour; and now, as it has done its work, its foundations are allowed to become soft, and to crumble away.[117]

Repairs apparently were not made, but portions of the wall remained for some time. In 1869, portions are mentioned as still standing, although in a dilapidated condition.[118] As late as 1875, when a survey was made to reconcile the Salt Lake townsite with the governmental survey, portions of the wall still standing were sketched in on the map appearing here as Figure 12.

OGDEN

Ogden, like other Mormon cities, also had a wall. A beginning was made in 1854. When Wilford Woodruff visited Ogden in December of that year, he reported: "The city wall will enclose one mile square and is to be built of earth, eight feet high, three feet wide at the bottom and eighteen inches at the top: but very little is yet built."[119] The wall had stone foundations and was to have four gates, one in each direction.[120]

When William Chandless visited the village

[115] "Eleventh General Epistle of the First Presidency," *Journal History of the Church*, April 10, 1854, p. 1.
[116] Jules Remy, *A Journey to Great Salt Lake City* (London: Jeffs, MDCCCLXI), Vol. 1, p. 190. Entry for Sept. 25, 1855.
[117] Richard F. Burton, *The City of the Saints* (New York: Harper, 1862), pp. 197–98.
[118] *Salt Lake City Directory and Business Guide for 1869*, p. 61.
[119] Article by Wilford Woodruff, *Deseret News*, Dec. 11, 1854.
[120] Ordinance of City Council, Feb. 17, 1855. Milton R. Hunter, *Beneath Ben Lomond's Peak* (Salt Lake City: Deseret News Press, 1944).

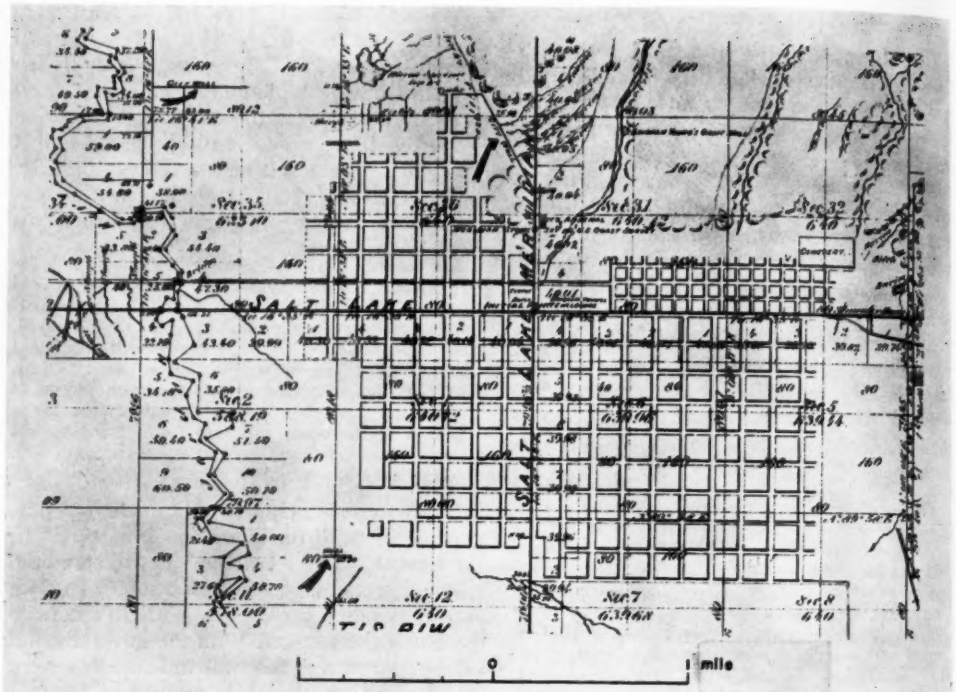


FIG. 12. Salt Lake City, 1875. A portion of the map of Salt Lake City was prepared by the Surveyor General to reconcile the original townsite of 1851 with the Federal Land Survey. Remnants of the wall were noted by the surveyor and on this print these remnants have been emphasized by arrows. Generally speaking, the entire area surveyed into blocks, both large and small (with the exception of the easternmost row of three lots south of the cemetery), was enclosed by the earlier wall.

in the winter of 1855–56 he found: "Ogden City was a specimen of the settlements in Utah on the model of Salt Lake; precisely a mile square, part on the bench, part in the valley-bottom, enclosed in an earthen wall, and laid out in blocks. . . ."[121] Presumably it, too, lasted only a few years.

MATERIAL AND DESIGN

American city walls at first glance may appear to have been frail and crude as compared with the walls of cities in Europe and Asia with which we are familiar. In the United States the majority of walls were constructed from wooden pickets, from eight to fifteen feet high, perhaps 10 inches thick, imbedded in the ground adjacent to each other, occasionally in double rows. Some of the walls were embellished with blockhouses, bastions, walkways, and so on, and some of the gates had gatehouses. Moats, dry or filled with water and provided with drawbridges, appeared in some instances. Occasionally portions of a wall, particularly the bastions, were made of stone, interrupting the pickets for a short distance. Exceptions to the general "picket" pattern were found in a few instances. Boston's short wall was, in its final form, stone. Savannah had walls of "meer sand" supported by "saplings" on both sides, and Salt Lake City and Ogden both had walls of packed earth reinforced with imbedded vegetation.

Thus many of the American city walls seem to have been a good deal less substantial than city walls found in other areas. Perhaps they were not even in the same category—more

[121] William Chandless, *A Visit to Salt Lake* (London: Smith, Elder and Co., 1857), p. 232.

stockades than walls. This feeling, however, may be due to the fact that the city walls that have been preserved elsewhere, or the walls that have been removed in the recent past, were the final, much-improved versions of earlier, less pretentious walls. For if we inquire further we find that many English medieval towns had defenses "if only a town ditch and stockade. A few had earth ramparts . . . [sometimes] stone walls were provided for the large towns."[122] Similarly, in China, where, until recently, the massive walls of Peking remained for observation, "the earliest towns in forested areas appear to have been surrounded with log palisades." Further, "in South East Asia wood or bamboo palisades were common."[123]

Although a good deal of criticism was directed against the design of American city walls, many of them were planned by military engineers or army officers who were presumably well trained in the art of urban fortification and were familiar with the walled cities in Europe. De Brahm, for example, who drew the plan for the walls of Savannah (and another set for Charleston which was not accepted because it was in French), was a military engineer.

EFFECTIVENESS

Walls surrounding American cities seem to have been effective in protecting the cities involved. None of the walled cities was ever stormed. Perhaps most of the value of the wall was in its deterrent effect, a comment that possibly could be made of city walls everywhere. It is true that Detroit was burned several times. Schenectady, too, although walled, was burned, but here the gate was left open and unguarded, and the invaders came in the gate and not over the wall. Cities that were open to the sea occasionally had to pay tribute to pirates, but in no instance do these city walls seem to have been scaled or breached.

Besides protection, the walls of cities may have provided other, less tangible, advantages. In New York, for example, when the gate was locked for the night, and in other cities when the gate was closed, a sense of civic belonging may have been generated, similar to that described by Mumford as one of the advantages of the Medieval city.[124] In New Orleans, it was asserted that the walls may have been as useful to impress the inhabitants within with Spanish authority as to discourage the enemy without.

SIGNIFICANCE

Robert Dickinson, in speaking of the relation of the wall to the present "townscape" of European cities, emphasizes that the "lie of the streets and the arrangement of the blocks show close adjustment to the wall, even when it has disappeared."[125] Could a similar comment be made about the American cities that have had walls?

The French word "boulevard" originally meant the horizontal portion of a rampart, and eventually the promenade, usually tree lined, laid out on the space made available by a demolished fortification. It is a common feature of many European cities. Rampart Street in New Orleans and Oglethorpe Avenue (formerly South Broad Street) in Savannah fit these specifications. Both seem to have been laid out on the open space available inside the walls, both are unusually wide, and both have their traffic lanes divided by a grassy, tree-studded parkway—classic boulevard features. Esplanade Avenue, a wide promenade, and the paralleling Barracks Street in New Orleans also seem to have been carved out of land formerly part of the fortified area along the south side of the early city.

The well-known Wall Street in New York City developed as a lane running just inside of that city's wall. Similarly, Wall Street in Salt Lake City, an unusual diagonal at the edge of the city's all-prevailing grid, marked the location of the earthen walls of that city. Today several neighboring streets have also taken on this diagonal alignment. Wall Avenue in Ogden also marks the line of the west wall of that city. In Charleston, the southern line of the wall followed a creek cutting diagonally to the Ashley River; today Water Street follows the former wall and creek. Similarly, Cumberland Street on the north runs at a slight diagonal, as did the earlier wall which it closely parallels.

In like manner, bastions or strong points in

[122] Sir William Savage, *The Making of Our Towns* (London: Eyre and Spottiswoode, London), p. 107.
[123] Joseph E. Spencer, *op. cit.*, p. 139.
[124] Lewis Mumford, *op. cit.*, p. 54.
[125] Robert Dickinson, *op. cit.*, pp. 271–2.

the wall-defense system occupying a wider area than the wall itself, occasionally remain as open places, or as some other form of public land. A fine example is Beauregard Square (formerly Congo Square) in New Orleans, earlier the location of the bastion Fort San Fernando. In the same city, the United States Customs House approximates the site of Fort St. Louis, which was located at the northwest corner of the wall. Similarly, Fort San Carlos, on the southwest corner, the site of an earlier United States mint, is now a Coast Guard supply depot. In Charleston, too, the former Crane's Bastion on the northeast corner of the wall had the same location as the present customs house.

Major streets in none of our walled cities carry the name of their destination, such as East Gate or Water Gate, and assessing the influence of the city gates is difficult. Gates stood astride New York's Broadway, for example, but it had been first an Indian trail and already was a major road before the wall was built. Here, as in other cases, the gate may have strengthened the street's development rather than originating its importance. Outside of the wall in New Orleans, Bayou Road leads away from a rear gate, a diagonal street then as now in the surrounding grid pattern. Castellanas, writing in the beginning of the twentieth century, indicates that perhaps memories of the gates lingered on for a while even if they have been forgotten today. He recalls, "The gate leading to the Bayou Road overlooked a drawbridge which spanned a wide and deep ditch at that point. In my boyhood years the place was still called *la porte du bayou*."[126]

In observing the present land-use pattern, one can occasionally sense the transition from within the wall to outside of it. In Savannah, a large old cemetery, now "The Colonial Park," is located just outside the line of the former wall (where it was reached through the "Burying Ground Gate"). In New Orleans, too, St. Louis Cemetery lies just outside the old line. In both Salt Lake and Ogden sizeable plots of land lying just outside the former walled area appear today as city parks. In at least three of the above examples, however, the wall surrounded the original plat of the city which preceded the construction of the wall. It is a possibility that cemeteries and parks were located so as to be outside the first plat rather than with reference to the wall.

Of these eleven cities only two, New Orleans and Schenectady, seem fully conscious of the location and circumstance of their former walls. New Orleans, of course, is famous for its "Vieux Carre," or old square (really an oblong), more popularly its "French Quarter." It is carefully and deliberately preserving the architecture and the spirit of the old city: all new construction and repairs in the area must be approved by the Vieux Carre Commission. Some of the maps sold in the city show the former line of the ramparts prominently and with reasonable accuracy.

In Schenectady, where the historic city was of a more modest scale, the "old stockade," as the earlier walled area is known, has remained residential as the business district migrated east on State Street. It seems almost possible to trace the line of the palisades through the back yards on the water sides, and a small incline to the south, indicated on early maps, is also visible today. Recently a Stockade Association dedicated to preserving the character of the old neighborhood has been formed. For some time it has been socially desirable to live in the "old stockade" and property values there are higher than for equivalent houses elsewhere. Many markers indicate the position of the fort and other old buildings. An apartment house along the line of the fortifications is called "The Stockade" and a gift shop in the same neighborhood in recent years has also carried that name.[127]

While it is true that the majority of the cities in the United States were never walled and thus developed free from their influence, a few cities were walled. And in some of the American cities that did have walls, these ancient and transitory barriers have influenced the development of the "townscape" in much the same manner, if to a generally lesser extent, as city walls have affected the urban landscape of cities in other areas of the world.

---

[126] Henry A. Castellanas, *New Orleans as It Was* (New Orleans: Graham, 1905), p. 316.

[127] Lewis B. Sebring, Jr., Associate City Historian, with whom I explored the "old stockade," furnished some of the information in the above paragraph.

SOME ASPECTS OF THE URBAN GEOGRAPHY OF THE CHINESE HSIEN CAPITAL[1]

SEN-DOU CHANG
University of Washington

FOR about 2,500 years before the Western impact on Chinese society in the middle of the nineteenth century, almost all of the urban population in China was concentrated in walled cities. Fortification was so essential to the Chinese idea of a city that the words for "city" and "wall" are identical; "ch'eng" signifies both a city and the wall of a city. The walled cities were essentially administrative in function, and the urban hierarchy in dynastic China paralleled the administrative hierarchy. The imperial capital, usually the most populous urban center in the nation, served as the government headquarters, the economic focus, and the cultural metropolis as well as the communications center for the whole empire. The provincial capital was usually the most populous city in each province. The hsien capital was the lowest administrative center through which the imperial government exercised its authority over the masses in the rural communities. The political status of the hsien capital is approximately equivalent to that of the county seat in the United States, although the two differ in certain functions. Before 1949 there were about 1,950 hsien capitals in the whole of China, compared to a small number of provincial capitals, which have never exceeded thirty through all the dynasties. The hsien capital, in a regional sense, was a miniature of the imperial capital, because it not only was the administrative center for rural communities but also served as the economic focus, the cultural center, the transportation node, and the central market, and, in most cases, was the most populous settlement of each hsien. Mainly because it contained a great majority of the total Chinese urban population and had such multiple functions in the rural areas, the hsien capital constituted an important element of the fabric of Chinese society.

Western literature on China includes much material on the nature and character of imperial capitals, treaty ports, and rural markets and villages. The hsien capital, however, has been relatively neglected. This article is an effort partially to fill that gap. It attempts, by discussing some historical, geographical, and morphological aspects of the hsien capital, to reveal some characteristic features of Chinese urbanization within the framework of the traditional culture which existed before Western culture exercised its influence on the coastal cities, and, of course, before the present drastic changes in the urban landscape had taken place under the Communist regime.

The article concentrates mainly on four important aspects of the hsien capital: its historical development, its site and distribution, its internal structure, and its major functions. Since the hsien capital was originally established for administrative convenience and for tax collection, a study of its spread and distribution can throw important light on population growth, frontier expansion, sequent occupance, and the regional economic development of China. As the formation of hsien capitals was influenced by many cultural factors, a structural study of them may reveal something about the physical and social basis of their sites, and about the ideological factors which influenced the internal arrangement and land-use pattern within the city. Finally, as the hsien capital grew out of a locally self-sufficient agricultural economy, a study of its functions may throw light on the nature and extent of its "sphere of influence"[2] and on urban–rural relationships in China. This may be helpful in testing the validity of some theories on urban geography in a wider scope.

A few findings of this article may be sum-

[1] Sincere appreciation is expressed to Professor Rhoads Murphey, Department of Geography, University of Washington, for his helpful criticism of the first draft of the manuscript of this article, and to Professor Hellmut Wilhelm, Far Eastern and Russian Institute, University of Washington, for his valuable suggestions on source materials used in this article.

[2] The area surrounding a town and functionally linked with the town has been denoted in different terms by geographers. Among the most common ones are "hinterland," "umland," "urban field," and "sphere of influence." To me, "sphere of influence" seems the most dynamic term.

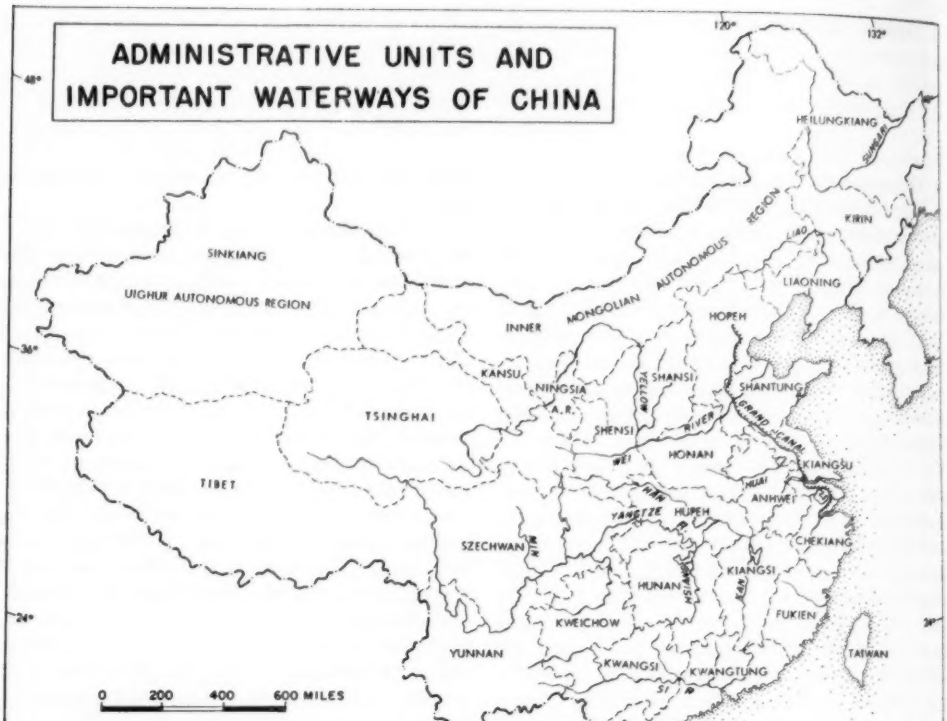


FIG. 1. Administrative units and important waterways of China.

marized as follows: (1) Except in a few cases, the hsien capitals were situated in lowland areas, especially along waterways. This location seems in contrast with that of some ancient or medieval cities in Europe, many of which were located on elevated or protected sites. This difference may reflect the dissimilarity of defense concepts and social structures in the two cultures. (2) Ancient China was essentially a "continental empire," since the establishment of large permanent urban settlements along seashores did not take place until relatively recent times. (3) The internal structure and land-use pattern of the walled city were strongly influenced by the ideological cast of traditional Chinese society, the distribution of temples, and the location of city gates. (4) The hsien capital functioned as the focus of a nodal region—the hsien (or county) — and occupied the most important position in the "central-place" hierarchy of China. (5) Many treaty ports and large urban settlements in China were hsien capitals originally.

EARLIER CHINESE CITIES AND THE ORIGIN OF THE HSIEN CAPITAL

According to available archeological evidence, urban settlements in China existed as early as the second millennium B.C., when bronze-using groups in North China were living in clustered settlements, composed of rectangular timbered houses forming quadrangular towns, defended by ramparts of tamped earth. The Shang people of the Bronze Age civilization spread along the middle course of the Yellow River and occupied a number of small, scattered, autonomous city states which existed like islands in a sea of Neolithic people. Shang society consisted mainly of two groups, the urban aristocrats and the rural peasants. This sharp distinction between ur-

ban and rural was still in existence during the feudal period of the early Chou dynasty.[3] As frontier expansion proceeded, cities were created for administrative purposes in newly captured territories. For protection against tribal invasions and for safe grain storage the construction of walls around the cities was necessary. The vassals in the city enjoyed a fiefholder's right to have natives in the country work on their lands. However, the feudal relationship of ancient China, in contrast to that of Europe, was relatively lacking in mutuality of rights and obligations, and was almost purely an economic relationship.[4] This gradually took the form of landlord-tenant relations. Later on, a group of tax collectors developed. The lords regarded these tax collectors as their "officials" and shaped their relations with them after the model of the household administration.[5]

Throughout the later part of the Chou dynasty, a period of frequent disorders and declining central authority, wars of conquest for the satisfaction of the personal interests of the lords became widespread. Powerful lords gradually made the point that conquest was the consequence of their own personal merit and thus they felt no obligation to delegate sovereign rights in new territories to any of their bondsmen. The lord created "military administration" in a walled city under the administration of an able tax collector and administrator. These centralized administrative centers in the frontier regions were most likely the origins of hsien capitals.[6]

[3] The question of whether it is justified to make use of the term "feudal" to describe the Chou society of China has been frequently studied and in general has been accepted. Granet points to significant institutional parallels between the society of the Chou dynasty and medieval Europe. Basic to both societies, in his opinion, is the bifurcation between an elite warrior class on the one hand, living according to an elaborate code of honor and having as its main function the performance of military service on behalf of the suzerains by whom it is invested with landed estates, and, on the other hand, a mass of despised peasants who lack this code of honor, play only a subordinate role in warfare as conscripted foot soldiers, and possess no right to the land which they cultivate. See Marcel Granet, *La Féodalité Chinoise* (Oslo, 1952). See, also, Rushton Coulborn, *Feudalism in History* (Princeton University Press, 1956).

[4] Wolfram Eberhard, *Conquerors and Rulers* (Leiden: Netherlands, 1952), p. 10.

[5] *Ibid.*, p. 11.

[6] *Ibid.*, p. 12.

The state of Ch'in was probably the first to establish modern hsien capitals. In 688 and 687 B.C., four newly acquired territories were converted into the first hsien of which we have record.[7] The same process also occurred in other states during the following decades.[8] In this stage of development the hsien system probably existed parallel to, but not in conflict with, feudalism, since it was applied only to new land that had been captured from non-Chinese tribes or from other Chinese states. Later, the hsien was introduced by various states into the territories of their own feudal vassals, especially in cases where a vassal had died leaving no posterity, or was becoming too powerful for the state's safety. An instance of the latter occurred in the year 514 B.C., when the lands of two powerful clans in the state of Ch'in were seized by the minister of that state and converted respectively into seven and three hsien.[9] It is evident that the institution of the hsien marked a trend away from pure feudalism and toward a more direct system of centralized state control.

This transition had some impact on the fundamental function of walled cities. The walled cities gradually shifted from being the private property of vassal families to regional administrative devices, operated by a representative official of the central government. Through the continuous influence of centripetal and centrifugal forces[10] on their surrounding areas, the walled hsien capitals gradually became the nodes of their small administrative regions.

THE SPREAD OF HSIEN CAPITALS

Throughout the time of the Warring States (approximately 400–221 B.C.), however, the establishment of hsien capitals was an exception rather than a rule. It was not until 221 B.C. that hsien capitals were uniformly established throughout the newly unified country under the reign of Ch'in Shih Huang, founder of the Ch'in dynasty, the first to establish an

[7] Derk Bodde, *China's First Unifier* (Leiden: Holland, 1938), p. 136.

[8] *Ibid.*, pp. 239–46.

[9] *Ibid.*

[10] For centrifugal and centripetal forces of a city acting on its surrounding areas, see Charles C. Colby, "Centrifugal and Centripetal Forces in Urban Geography," *Annals*, Association of American Geographers, Vol. XXIII, No. 1 (March, 1933).

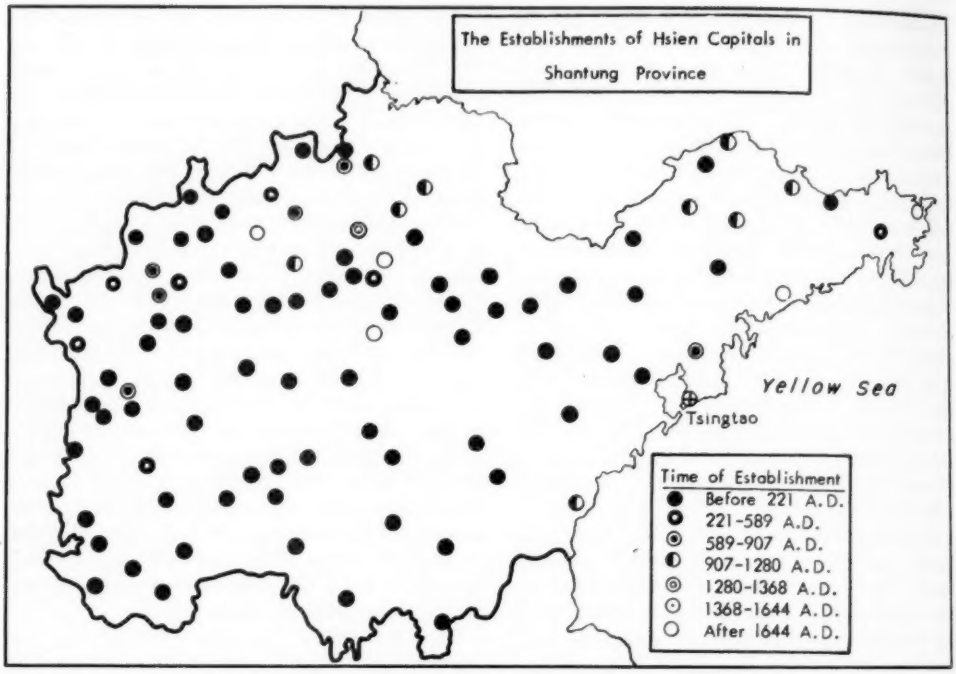


FIG. 2. Time of establishment of hsien capitals in Shantung Province. (Source: *Shantung T'ung Chih*, Shanghai, 1934.)

effective central administration. Its successor, the Han dynasty (206 B.C.–222 A.D.), noted for its enormous territorial expansion, saw the greatest number of hsien capitals established. For instance, among the ninety-two hsien capitals of Shensi province, the most important economic area of ancient China, thirty-six were founded during the Han.[11] Seventy-six out of a hundred and seven hsien capitals in Shantung province, the eastern fringe of the North China plain, were also established during this period (Fig. 2). The southward advance of colonization under the Han seems to have followed the two important tributaries of the Yangtze, the Hsiang and the Kan, across the Nanling Mountains, and to have reached the lower portion of the Si River around Canton. Twenty-two hsien capitals already existed in the territory of the present Kwangtung province before 222 A.D.[12] In 67 A.D., even on Hainan Island there was a Chinese community of considerable size settled in the walled city of Chuyai ("pearl shore," near the present city of Kiungchow) among the many tribal settlements of Li people.[13] In contrast to the early development of the Si River region in the extreme south, the hill region along the southeastern coast was not significantly occupied by Chinese until about five centuries later, in the T'ang dynasty (618–907 A.D.). The fact that only fourteen of the sixty-six hsien capitals of Fukien province were established before the seventh century tends to support this assertion.[14]

Unlike the rest of the country, the southwest region was relatively neglected by the Chinese of the north, and Yunnan province is probably among the latest territories settled by the Chinese in China proper. Among the 120 hsien capitals of the province today, there

[11] *Shensi T'ung Chih*, Vol. I, Chap. 5.
[12] *Kwangtung T'ung Chih*, Vol. I, Chap. 3.
[13] William Frederick Mayers, "A Historical and Statistical Sketch of the Island of Hainan," *Journal of the North China Branch of the Royal Asiatic Society*, 1871.
[14] *Fukien T'ung Chih*, Vol. I, Chaps. 5–6.

was not a single one permanently established before 1270 A.D.,[15] in spite of the fact that Chinese military campaigns repeatedly reached this territory during the Three Kingdoms period (221–265 A.D.) and in the T'ang dynasty. The Yuan (1279–1368 A.D.), one of the shortest dynasties, created few hsien capitals in China proper. However, this powerful empire, with its immense territorial expansion in all directions, seems to have contributed a great deal to the development of the southwest corner of China proper. During the reign of Kublai Khan alone, forty-five hsien capitals were erected in Yunnan province.[16] The opening of the Yunnan plateau for large-scale Chinese colonization was associated with the great military campaign against the Thai kingdom of Nan-chao and against native tribes in the southwestern territories under the leadership of this vigorous emperor. In 1277, military forces of the Yuan advanced as far as the middle course of the Irrawaddy.[17] The second boom of town development in Yunnan occurred during the most flourishing years of the Ch'ing dynasty (1644–1911). In the four-year period from 1727 to 1731, another ten hsien capitals were added to the landscape of the Yunnan plateau.[18] This is related to an important event affecting the status of the aborigines there under the reign of the Yung Cheng emperor. The Miao people in certain parts of that province had been permitted for ages to retain their native institutions, by testifying to their allegiance to the emperor through paying periodic tributes. Due to their constant refusal to fulfill such obligations, Ortai, the Manchu viceroy of Yunkuei, decided in 1726 to bring them under the direct rule of the government.[19] Guerrilla warfare occupied the government there for four years, at the end of which the viceroy was able to convert these territories into hsien. The rest of the hsien capitals in that province were products of the nineteenth century.

By approximately the time of the downfall of the Yuan empire, Chinese colonization in China had been largely accomplished in every corner of the so-called Eighteen Provinces,[20] each being dotted with walled cities. The creation of hsien capitals in the Ming (1368–1644) and Ch'ing dynasties was mainly of two kinds: (1) new hsien were added by splitting old hsien which, due to population increase, needed additional hsien capitals for convenience in administration and taxation; (2) new hsien capitals emerged in the coastal, river, and delta regions in South China. The orientation of Chinese settlement gradually shifted toward the sea. According to Ta Chen's estimate, which was based mainly on official historical records, the Chinese population showed little change during the first seventeen centuries of the Christian era (roughly from the Han dynasty to the end of the Ming dynasty). Most of the estimates were between 50 and 60 million.[21] But from the middle of the seventeenth century to the middle of the nineteenth century, the population showed a rapid increase. Ho Ping-ti's compilation of official population data from 1741 to 1850 shows the following figures:[22]

Year	*Population*
1741	143,411,559
1750	179,538,540
1775	264,561,355
1800	295,273,311
1825	379,885,340
1850	429,931,034

Since most new hsien capitals of this period were established between old hsien capitals in the river deltas and flood plains of South China, there is reason to believe that the rapid population increase of this period was not due to territorial expansion or frontier colonization, but rather due to intensified use of pre-

[15] *Yunnan T'ung Chih*, Vol. II, pp. 2–4.
[16] *Ibid.*, Vol. III.
[17] Henry H. Howorth, *History of the Mongols*, Part I, "The Mongols Proper and the Kalmuks" (London, 1876), p. 242.
[18] *Yunnan T'ung Chih, op. cit.*, Vol. III.
[19] *Ch'ing Shih Kao*, T'u Ssu Ch'uan 3, Yunnan.

[20] The term China Proper is sometimes used for the classic eighteen provinces south of the Great Wall and east of Tibet.

[21] Ta Chen's estimates for China's population in history were based on classic historical records and show the following figures:

Period	*Population*
2 A.D.	59,500,000
742–756 A.D.	52,900,000
1098–1100 A.D.	43,000,000
1573–1620 A.D.	60,000,000

See Ta Chen, "Population in Modern China," *American Journal of Sociology*, July, 1946.

[22] Ho Ping-ti, *Studies on the Population of China 1368–1953* (Harvard University Press, 1959), pp. 281–82. Appendix I.

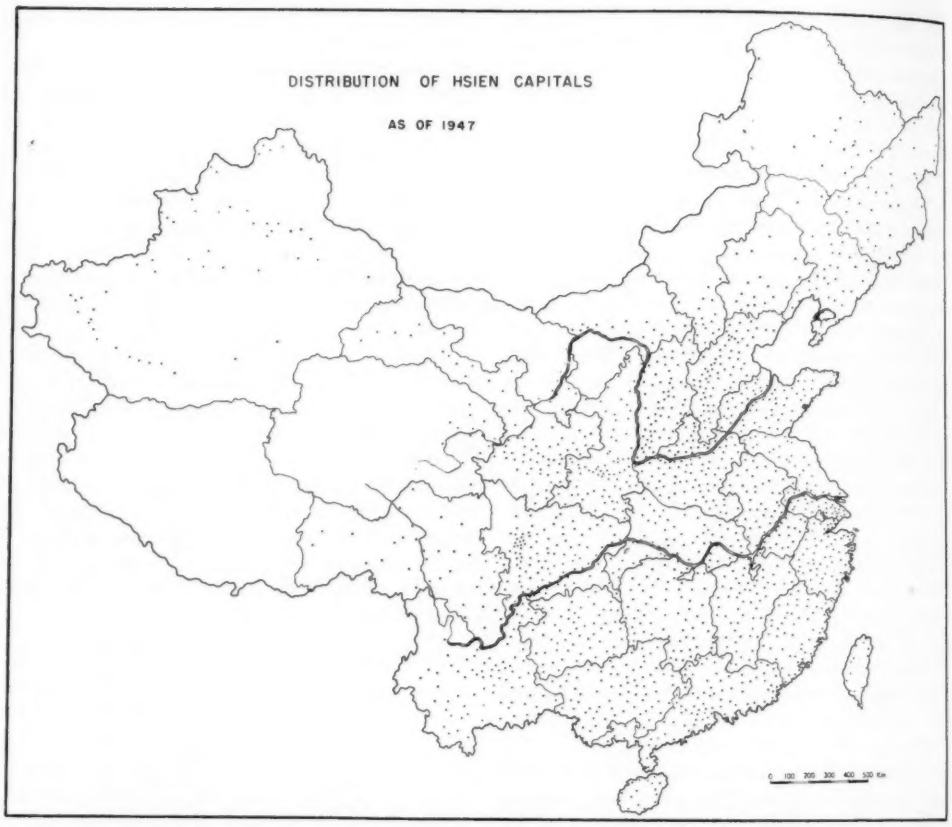


FIG. 3. Distribution of hsien capitals. (Source: *The Provincial Atlas of China*, Shanghai, 1947.)

viously developed productive areas, especially along the Yangtze and Canton deltas. It is probable, then, that the heavy concentration of the present Chinese population in these areas has formed or at least become apparent since the middle of the seventeenth century.

In the lower part of the Yangtze delta almost all hsien capitals were established after the tenth century. Eight hsien capitals near Shanghai were not established until the early part of the nineteenth century.[23] This suggests that the lower Yangtze delta was settled relatively recently. The area has emerged from the sea due to silt deposits carried by the Yangtze. According to V. K. Ting, the coast of the delta is pushed seaward one mile every 69 years.[24] He therefore believes that 5,000 years ago the coastline was near the present Kiang-yin hsien capital, which is now about eighty miles inland. He also believes that Chung-ming Island did not emerge until the middle of the eighth century. Five hundred years later, the island was large enough and populous enough to be established as a hsien.

THE DISTRIBUTION PATTERN OF HSIEN CAPITALS

In contrast to the Roman empire, which depended greatly upon sea traffic in the Mediterranean, ancient China was essentially a "continental empire," even though there is a coastline of more than 6,000 miles. Except for a few river deltas in South China, the sea

[23] *Kiangsu T'ung Chih*, Vol. I, Chap. 19.

[24] V. K. Ting, "Report on the Geology of the Yangtze Valley Below Wuhu," *Shanghai Harbor Investigations Series Report 44* (Wanpo Conservancy Board, 1919).

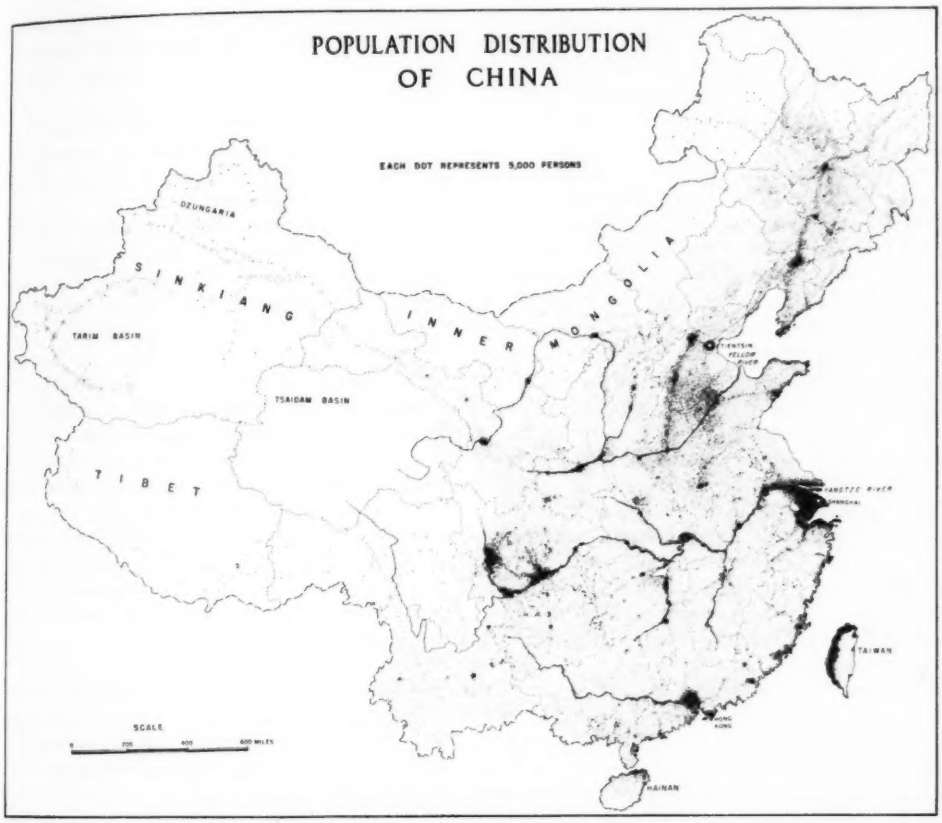


FIG. 4. Population distribution of China, 1953.

coast rarely attracted large communities or urban settlements until relatively recent times. Overseas trade as an agent of urban development was seldom important in China. While the Mediterranean Coast was embroidered with many great port cities, such as Alexandria, Marseilles, Syracuse, Tyre, Ostia, and Tripoli in the first century B.C., the coastal areas of North China remained largely unsettled. When the area of the present Shantung province was booming with walled cities in the Han dynasty, its coastal area was quite neglected. Among twelve hsien capitals east of the longitude of Ts'ingtao, only three were in existence during the dynasty (Fig. 2). The walled city of Yung Chen, situated at the tip of the peninsula, was founded as late as 1731.[25] Even at the present time, many coastal areas still lack urban settlements, especially along the coasts of Hopeh and northern Kiangsu (Fig. 3). The Grand Canal, among others, was built essentially for shipping grain along an inland waterway and consequently promoted the development of inland towns and reduced the importance of coastal towns. Though a few cities along the southeast coast, such as Ch'uanchow (Zayton) and Canton, had been trade centers of international significance from the seventh century, their number was small as compared with the Mediterranean area for the same period. Chinese relative avoidance of the sea became more apparent when the Western powers were searching for port cities on the China coast for trade or for military bases during the nineteenth century. Few cities with good natural harbors were in existence at that time. Many European countries had to create new port

[25] *Shantung T'ung Chih*, Vol. I, p. 881.

cities themselves on the sites of Chinese fishing villages. Ts'ingtao, Weihai, Dairen, Port Arthur, Hong Kong, among others, are all Western creations of the nineteenth century, lacking any historical background on the Chinese side.

In comparing the distribution map of Chinese population (Fig. 4) with the distribution map of hsien capitals for the last few decades (Fig. 3), one will find hsien capitals relatively more evenly distributed than the population. Carr-Sanders has attempted to explain the growth of towns and cities as a reflection of growth in the density of rural population in some underdeveloped countries.[26] Such an explanation would have validity in an area where the growth of towns is entirely due to regional economic development and the need for central markets for rural areas. In applying such a generalization to correlate the number of walled cities and the density of rural population in China, however, one must proceed with some reservation, since the closer spacing of walled cities in certain regions does not necessarily indicate a greater density of rural population. The divergence of these two phenomena may be explained by various factors. At least two of the factors are clear.

1. In the western portion of the loess region, the population density is moderate, but the density of hsien capitals is comparable to that of the middle Yangtze valley (Fig. 3). These hsien capitals in western Shansi and eastern Kansu were mainly garrison towns during the Han and T'ang dynasties and were established for protection against Mongolian and Turkish invasions. The number of residents in these walled cities was largely dependent upon the growth or reduction of garrison troops. After the seventeenth century when Mongolian and Turkish influences were gradually diminishing, these garrison towns lost most of their importance. Nevertheless, the walled cities remained, and a small local population lived on a subsistence basis. Perhaps these frontier hsien capitals are not much larger than the village settlements in the Yangtze delta area, but their administrative functions for their surrounding areas still exist. The names of these hsien capitals suggest their original function as garrison towns. For example, Chin Pien means "peaceful frontier," An Sai means "castle for peace," Si Ning means "peace in the west," and Yung Ning means "eternal peace."

2. Since the territorial expansion and frontier colonization of the Chinese were mainly accomplished by the time of the fall of the Yuan dynasty (about 1368 A.D.), a majority of the hsien capitals were in existence by that time. Nevertheless, the great population increase did not occur until the middle of the seventeenth century. The Ch'ing dynasty established few hsien capitals, compared with the gigantic size of population increase for the same period. Therefore, if the distribution pattern of hsien capitals can be considered an index of the pattern of population distribution, it would show a rough picture of population distribution for medieval China rather than that of contemporary China. By the same token, any study of the structure and dimensions of walled cities would suggest primarily the degree of economic development for the particular period in which the walled cities were established.

SITE AND INTERNAL STRUCTURE

One of the characteristic features of hsien capitals is their sites. Unlike many ancient European towns, Near Eastern towns, or even the daimyo's castle towns in Japan,[27] the walled cities in China are almost exclusively located in lowlands, with consistent avoidance of conspicuous relief as much as the local configuration will allow. Rarely is a town in the south sited on the slope of a mountain or a hill, and hardly ever in the north. Out of the sixty-six hsien capitals of Fukien province, a province with rugged topography, sixty-one are situated at less than 400 meters of altitude (Fig. 5), sixty of them being situated along rivers, streams, or other water bodies (Fig. 6). This lowland-oriented urban phenomenon is closely associated with the fact that the Chinese population has been overwhelmingly concentrated in lowland areas—flood plains, river valleys, intermontane basins, and small oases

[26] A. M. Carr-Sanders, *World Population* (Oxford, 1936).

[27] According to Glenn Trewartha, the daimyo's castle town occupied some sort of elevation, such as a hard-rock outlier of diluvian spur—a type of site that made it doubly conspicuous and more easily protected. See Glenn T. Trewartha, "Japanese Cities, Distribution and Morphology," *Geographical Review*, Vol. XXIV, No. 3 (July, 1934).

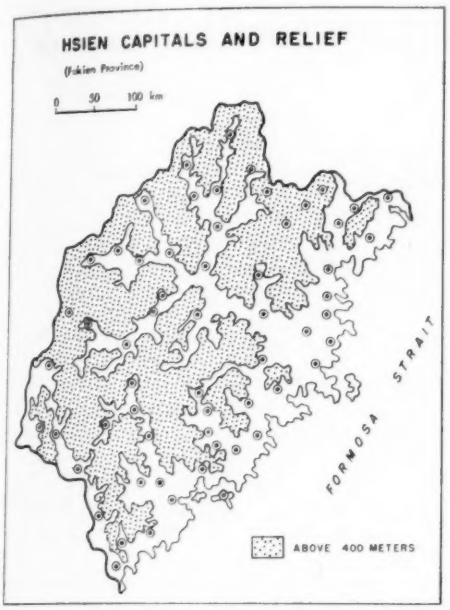


FIG. 5. Hsien capitals in relation to relief.

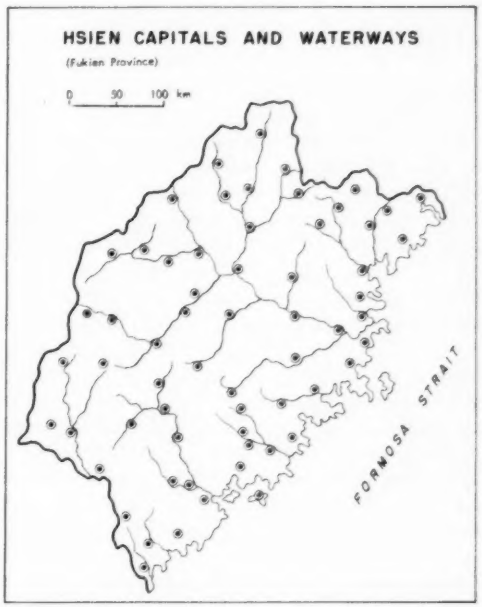


FIG. 6. Hsien capitals in relation to waterways.

along foothills, while upland areas have been largely unsettled and undeveloped, compared with those of many European countries.[28] Cities and towns arose in productive lowland areas and served the basic economic, political, and social needs of the rural population. Navigable waterways were an important means for connecting cities with rural communities or with other cities. Mining towns, lumbering centers, and resort towns, which are common features in many mountainous parts of the Western world, seldom existed in China.

As there was such a close relation between rivers and hsien capitals, most hsien capitals were "river towns" in character. The bank selection of hsien capitals along a river varies. For the nation as a whole, however, certain generalizations can be stated.

1. There is a distinctive difference in bank selection between the north and the south in lowland areas where land is equally productive on both sides of a river. In north and central China, most hsien capitals alternate regularly between opposite banks of a river, as exemplified by the Han River below Tzuyang (Fig. 7). This probably shows that there was a somewhat equal need for transshipping facilities on both sides. In many parts of South China, however, hsien capitals were located mainly on the north bank (Fig. 8). This phenomenon may be associated with the fact that during the course of the southward movement of Chinese colonization, walled cities established on the north bank of a river were comparatively safer than those on the south bank, since they were better protected against tribal attacks from south of the river.

2. In areas of rugged terrain, where a river divides productive land from less productive land, hsien capitals tended to line up on the bank with the productive land. The upper course of the Han River provides a good example (Fig. 7). The south bank is at the foot of a great barrier, the Ta Pa Mountains, between the Szechuan Basin and the loess region, and the north bank has access to the most prosperous economic region of ancient China, the Wei valley.

3. In an area where land transportation is replaced by waterways, hsien capitals tended

[28] The Chinese ignorance of upland use has been frequently discussed by Western scholars. See Pierre Gourou, "The Development of Upland Areas in China," in *The Development of Upland Areas in the Far East*, Vol. I, Part 1 (New York: Institute of Pacific Relations, 1949), pp. 11–13.

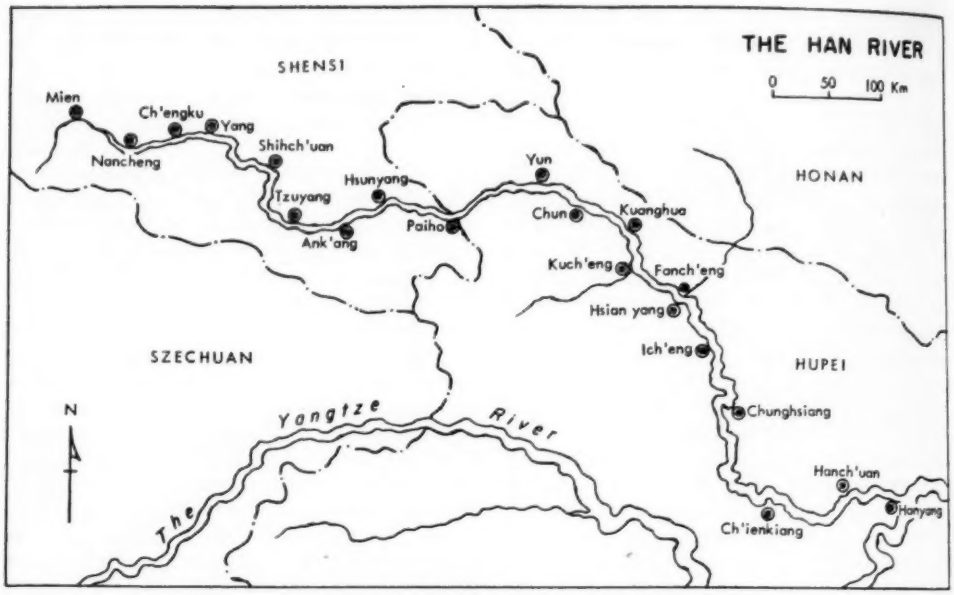


FIG. 7. Hsien capitals along the Han River.

to be located on the bank which received the largest volume of bulky commodities. For example, the west bank of the lower Min River in Szechuan province is more hilly than the east bank, since the former is near the eastern margin of the Tibetan Plateau. However, as the west bank is the lowland part of the basin nearest to the plateau, there has been a con-

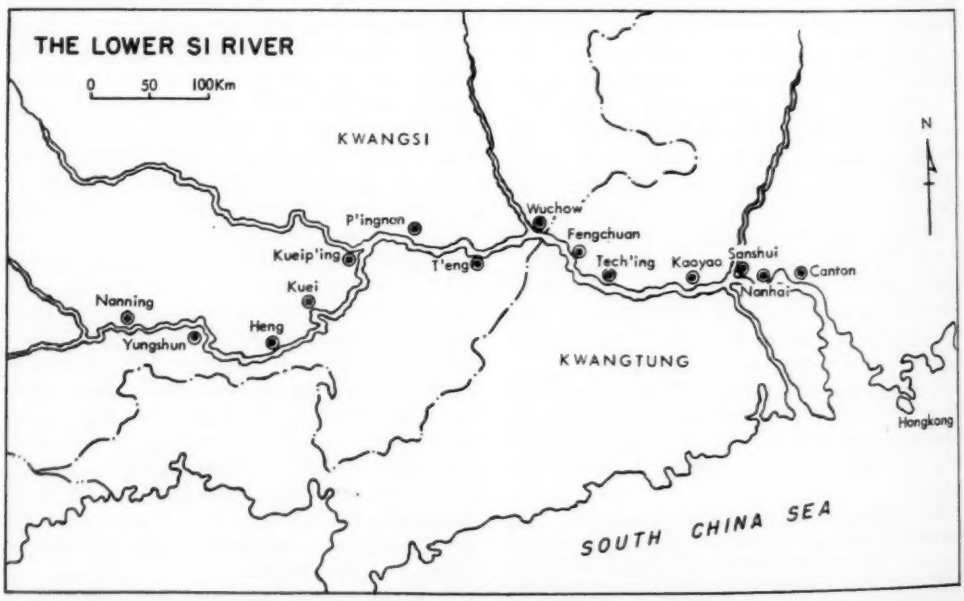


FIG. 8. Hsien capitals along the Lower Si River.

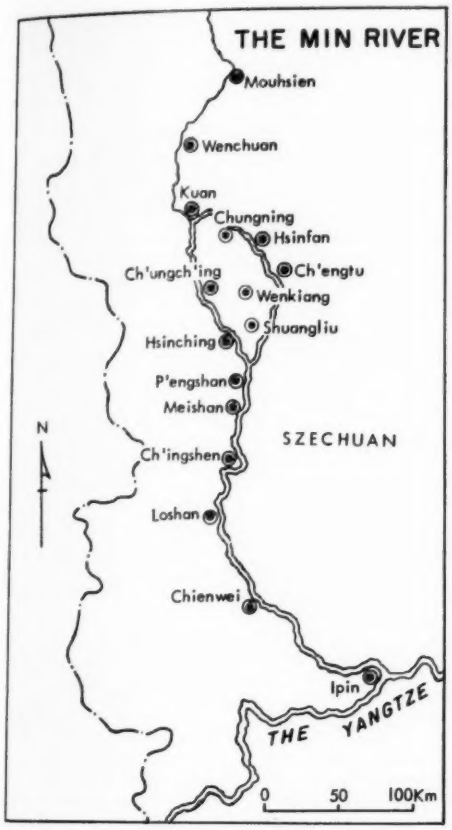


FIG. 9. Hsien capitals along the Min River.

siderable amount of exchange of native products between these two contrasting geographic regions from the time of the early Ch'ing dynasty. Tibetan products such as herb drugs, hides, precious minerals, and lumber usually came by land to the west bank of the Min for water transportation on junks or rafts. This "break-of-bulk" advantage probably constituted one of the main factors responsible for the rise of hsien capitals on the west bank of the Min River below Hsinching (Fig. 9).

Military considerations that would lead to choosing a site on a mountain ridge or hilltop were seldom of major importance in China. This is probably connected with some basic concepts about tactics of warfare. The whole area around a town was considered to be a defensive unit, and the protecting positions were on the margins of this area rather than the town itself. Therefore, the walled city of Lo-yang, a lowland city which was chosen for the site of the imperial capital many times in the course of Chinese history, was considered to be one of the most strategic points in China by many historians, because it is encircled by four defensible physical features: the Gate of Tiger Cage, about 50 miles to the east; the Gate of Fathomless Vale, about 90 miles to the west; the mountain of Prostrate Buffalo, about 70 miles to the south; and the Yellow River, roaring through the northern outskirts. These four features constitute a naturally protected zone difficult of access.

Such a concept of city defense is in contrast with that of medieval Europe, where the castle at the center of a town or a cliff at the edge of a city was considered to be an effective means of military defense; in other words, the medieval cities in Europe in many cases protected only themselves, without considering the surrounding rural areas. Such a contrast probably reflects some structural differences between the two societies.

First, the walled cities in China were essentially administrative in function. Economically they were the product of an agricultural society. The gentry group in the walled city was largely dependent on its lands and properties in the country. There was a strong linkage between the rural and the urban in China. It was the city dweller's duty and in his interest to protect rural areas in time of warfare. On the other hand, the medieval towns in Europe, according to Pirenne's study, were founded mainly to meet the needs of trade.[29] The middle class in the city formed a new, independent, self-governed society which paralleled the agrarian communities in rural areas. There was a closer economic tie between town and town than between town and country. Many towns even had to establish their own defensive units to protect their rights and profits against the conservative countryside. Thus a great number of towns in Europe were established on hilltops or cliffs overlooking the surrounding area.

Secondly, the Chinese hsien capital, functioning for the central government, was under the protection of imperial troops. In the event of barbarian invasion or local tribal rebellion, imperial troops stationed all over the country

[29] Henri Pirenne, *Medieval Cities* (Princeton, 1925).

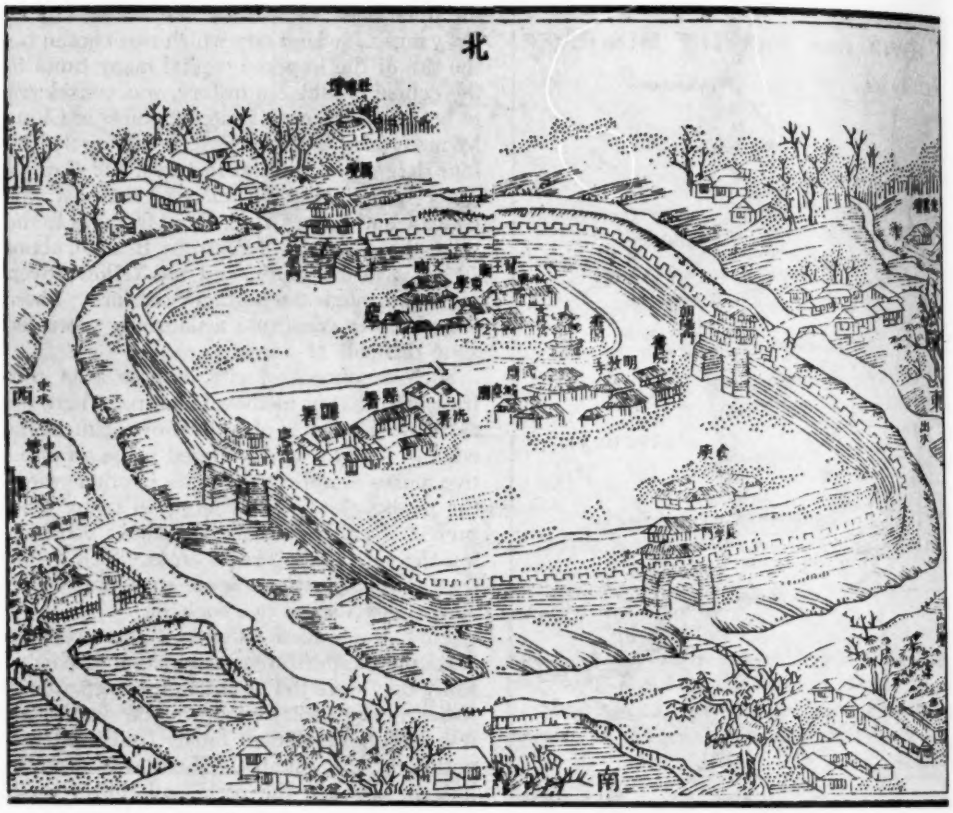


FIG. 10. A walled city surrounded by moats. (Source: *Chin T'ang Hsien Chih*, 1810.)

could be mobilized to concentrate in the troublesome locality. Thus, there was less need to build walled cities on the highlands. On the other hand, many cities in medieval Europe were independent not only in political status, but also in their economic activities and military functions, and in many cases they could hardly maintain an army of sufficient size to protect the city effectively. They had to rely heavily on prominent physical barriers for defense.

Thirdly, the major wars of China were fought mainly in the "marginal territories"[30] of Inner Mongolia and the northern portion of the loess region. The so-called "China proper," especially south of the Yangtze, experienced fewer disorders than did Europe where, from Roman times, barbarian invasions, religious wars, and international conquests reached most parts of Western Europe. In China proper, banditry or local tribal rebellion could, in many cases, effectively be kept outside the city by the brick-built city wall and the moats around the wall (Fig. 10). The consideration of defensibility in selecting the site for a Chinese city was probably not as important as in Europe. The acropolis, the city set upon a hill, which epitomizes a familiar geographical association in the Mediterranean and other parts of the world, is largely absent among the walled cities of China.

The Chinese preference for a level site may also have been influenced by the tendency to plan city walls in accordance with the traditionally recognized layout, in squares and rectangles, so that the whole structure needed a more or less level terrain. In the hill regions

[30] The term "marginal territory" has been used by Lattimore. See Owen Lattimore, *Inner Asian Frontier of China* (American Geographical Society, 1951).

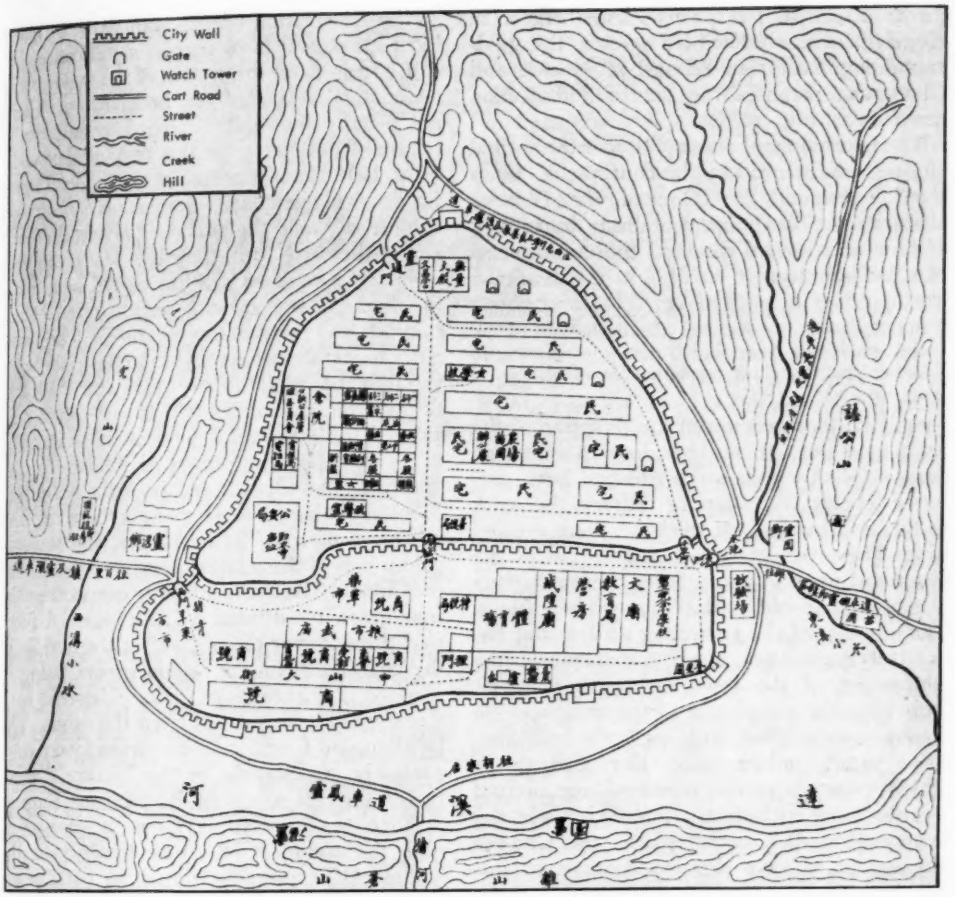


FIG. 11. The site of a walled city. (Source: *Lin Tai Hsien Chih*, 1935.)

where local configuration does not provide much level land for building cities, however, the pattern of the walls was so arranged as to be in perfect harmony with nearby elevated contour lines, at the expense of the square or rectangular shapes. The hsien capital of Lin-tai in Kansu province serves as an example (Fig. 11).

Water was probably the most important single factor influencing the site selection of hsien capitals, since it provided an important means both for transportation and for irrigation. As we can see in Figure 6, the most favored place for a hsien capital was the bank of a tributary or a main river. In the northwest, where the rainfall is less than ten inches, many oases served as the only possible spots for supporting settlements large enough to form hsien. Along the newly built Lanchow–Sinkiang railway, many old hsien capitals such as Kanchow, Suchow, Yumen, and Anhsi are located near alluvial fans, which are naturally irrigated every spring by snow water tumbling down from the Nanshan range. The association of hsien capitals with water bodies can best be illustrated by the literal meaning of their names, which often suggest site characteristics. The following are a few samples picked from various provinces: Ho-chuan means "river confluence" (in Szechuan province), San-shui means "three rivers" (in Kwangtung province), Chu-chi means "winding creek" (in Yunnan province), Kiang-kou means "the mouth of the river" (in Kweichow

province), P'ing-ku means "level vale" (in Hopeh province), King-chi means "the gold-containing river" (in Kiangsi province), and Chang-sha means "long sands" (in Hunan province).

The planning and construction of the walled city were influenced by the concepts of "feng-shui" (geomancy).[31] One of the Chinese classics, the *Shu Ching*, contains a description of how an overlord directed the construction of a walled city:

> To ascertain the points of the compass, he studies the shadows. He examines the declinities in sun and shade, the Yang and Yin of the country, to know how the chief constituencies of the world are divided.[32]

Finally he took account of the direction of the running water. It was his duty to select an auspicious site. When the site had been definitely chosen, the founder ordered the work to begin. The walls of the town were erected first. They were the most sacred part of the town and were in many cases the only structure which would outlast many generations. The erection of the encircling wall was of the greatest importance, since it determined the conception of the town as a single entity. Only after the completion of the walls was the marked space filled with separate buildings. As a result, unlike many European towns which grew up in the medieval age around a center—the cathedral, a fortified castle, the town hall, or the market place—the Chinese city seems to have developed more or less simultaneously on a plane defined by the walls.

The description of the wall gates in many local gazetteers is of special significance, because from them the system of the main streets connecting the gates can be deduced. There are usually four gates, placed symmetrically on the front sides of the rectangular walls (Fig. 10). Each gate has its own special significance. The east gate represents spring and brightness since the sun rises above the horizon of the earth from the east. The south gate symbolizes summer and life, as the wind from the south in the growing season often brings rainfall and warm temperatures. The west gate indicates autumn and sacredness, because through this gate the ancestor's spirit went to the western mountains which were often considered to be the residences of the gods. The north gate signifies winter and calamity, and was often closed, as it was from the north that Mongolian invaders frequently entered, and it was also from the north that the severely cold winds of the winter blew throughout most of China. The main streets, connecting the four gates, constituted a Latin cross. A fortified drum tower often stood in the middle at the crossing of the four main streets, being so situated that in times of strife or disorder each street would be isolated from the other. The location and function of the drum tower were probably comparable to the "piazza d'armi" in the fortress towns of Renaissance Europe.

Since towns were laid out in connection with the conquest of a new territory or the settlement of migrants from other parts of the country, the early hsien capitals were often garrisoned for protection against local tribes, and they also served as a place of refuge for the rural population of the district. Even today many hsien capitals still contain within their walls large stretches of cultivated land which were originally reserved to produce food for the population in case of siege (Fig. 12). As these farm lands are situated in an urban landscape, they have become an example of the most intensively used land in China. Since they are so close to the urban settlement, they are the most valuable farm lands in the whole hsien, and only high-priced cash crops, such as truck crops, fruits, or mulberry trees, are planted. Von Thünen's land-use pattern still holds true in many hsien capitals of China.

The main streets usually formed a clear

---

[31] Feng shui literally means "wind" and "water" in Chinese. This was a popular superstition connected with the worship of ancestors and of nature. To the Chinese all of nature was alive with influences for good or evil, as was revealed to those who made a study of their indications: the course of a stream, the trend of a mountain, the position of a clump of bamboo or a pagoda, the curve of a road, the site of a grave or a temple, the location of a settlement, and a number of other things—to the initiated eye all these formed visible manifestations of nature's future actions, or of the good and evil intentions of the departed dead. Thus, any construction work placed by man on the landscape must be made to harmonize perfectly with local physical features, by applying certain magic principles. For the origin and nature of this superstition, see James Dyer Ball, *Things Chinese* (London, 1900), p. 265; or S. J. Henry Dore, *Researches Into Chinese Superstitions*, Vol. IV (Shanghai, 1922), pp. 402–16.

[32] *Shu Ching*, "The Book of Chou," Section I.

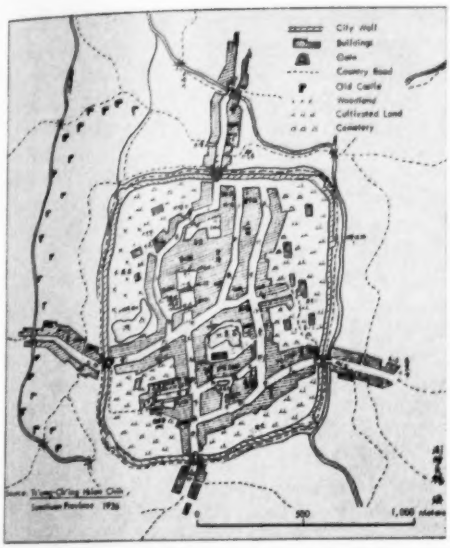


FIG. 12. The land use pattern of a walled city. (Source: *Ts'ung Ch'ing Hsien Chih*, 1926.)

pattern, as they connected the gates, but the secondary streets were often unsystematically laid out. Here houses were the primary elements and the layout of the streets was dependent on them. The entrances of houses were oriented to the south, so far as possible, in order to receive maximum sunshine in winter and sea breeze in summer. Streets were prevailingly narrow. They were seldom smooth, and paving was rare. Low, closely spaced, gloomy structures, often flush with the street, lined both sides. There were no sidewalks and no lawns. The most distinguished buildings were ordinarily the official residence of the magistrate and some temples for public pilgrimage. A few of these temples were common to nearly all hsien capitals; namely, the Confucian Temple, the Temple of the City God, and at least one local Buddhist temple. Woodlands of various sizes were often found immediately surrounding these temples. It was an ancient ritual to select trees from the finest groves to put in the sacred forests as an offering to the holy spirits. Many of these forests are still well preserved in hsien capitals, their appearance contrasting greatly with the countryside outside the walls, where many centuries of extensive deforestation have left the land comparatively treeless.

In the hsien capital, the business district was not nearly so evident as it is in the modern cities of the Western world. Many streets harbored both residential and commercial functions, for the dwelling and retail shop were in one building, the shop occupying the street side, the family living in the back or in the second-story rooms. This reflects two distinctive characteristics of a pre-industrial society: namely, the family system as a business unit, and the direct transactions between producer and buyer. In some large hsien capitals where convenient water transportation attracted greater commerce, different categories of craftsmen and shops were housed in different streets, a feature characterizing many trade centers of the world in the pre-industrial period. Sometimes the shopkeepers of a street formed a kind of guild, with a leader who represented their common interests, so that areal and professional organizations were identical. In the walled city of Chungking, for instance, the street names either denoted the most prominent structural phenomenon of the street or, in most cases, revealed the occupational function of that street. A few conspicuous examples are as follows: Shih Pan Kai means "stone-slab street," a street of solid cut stones; Chang Ping Kai, "long, level street," more level than most; Ta Tieh Kai, "iron pounder's street"; Ta Tung Kai, "copper beater's street"; Mien Hua Kai, "cotton street"; Mei Hua Kai, "plum-blossom street" (a flower market); and Yu Shih Kai, "oil-market street."[33] Country folk on their first shopping trips to the city found little difficulty in locating the proper stores by street names.

Though the hsien capital appeared to be the largest urban settlement in the hsien, it was essentially, in the economic sense, a product of a self-sufficient agricultural society. As such, its internal structure and its land-use pattern were not the same as in modern industrial or commercial centers. The internal structure of the hsien capital of the last few decades has been very much influenced by the location of ancient administrative and religious buildings as well as by accessibility to the gates. Near the Confucian Temple, which was usually as old as the city itself,

[33] J. E. Spencer, "Changing Chungking, the Rebuilding of an Old Chinese City," *Geographical Review*, Vol. XXXIII, No. 1 (January, 1943).

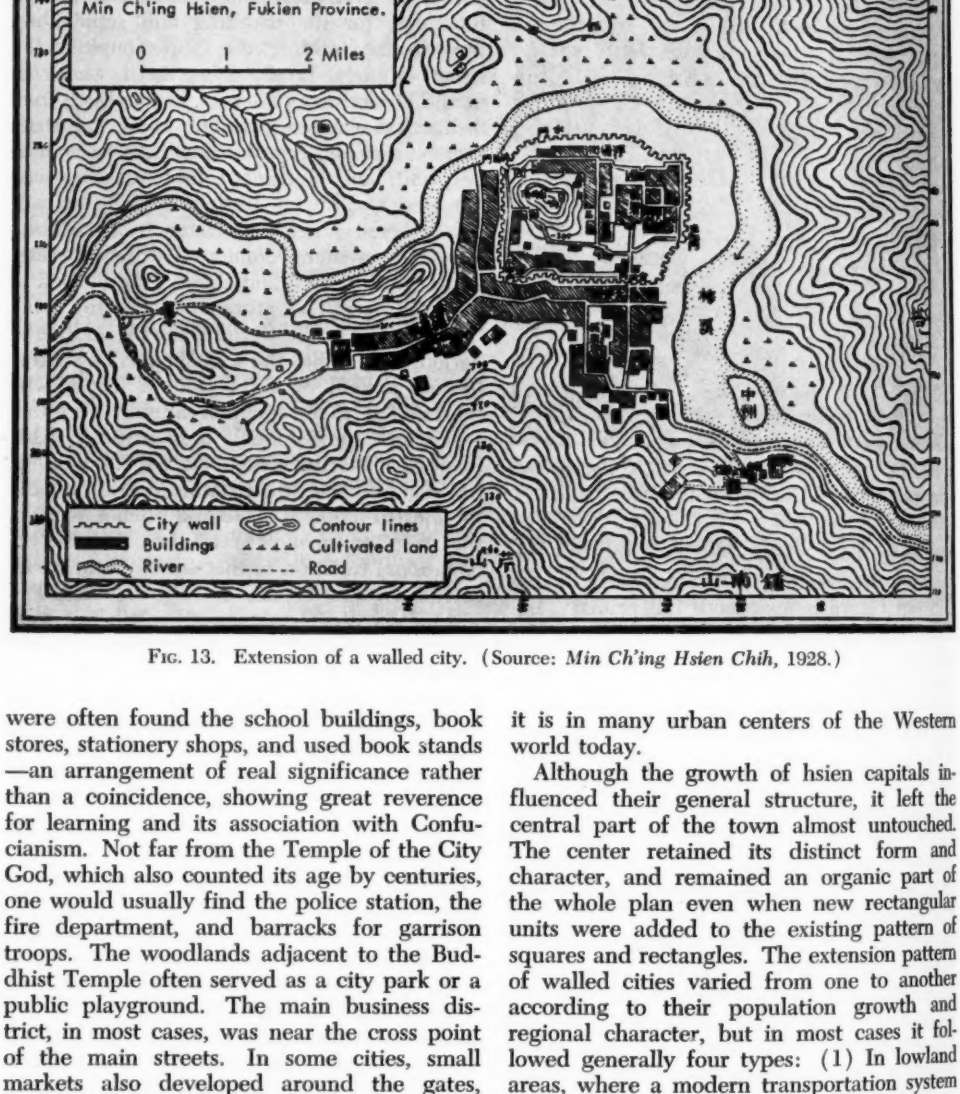


FIG. 13. Extension of a walled city. (Source: *Min Ch'ing Hsien Chih*, 1928.)

were often found the school buildings, book stores, stationery shops, and used book stands —an arrangement of real significance rather than a coincidence, showing great reverence for learning and its association with Confucianism. Not far from the Temple of the City God, which also counted its age by centuries, one would usually find the police station, the fire department, and barracks for garrison troops. The woodlands adjacent to the Buddhist Temple often served as a city park or a public playground. The main business district, in most cases, was near the cross point of the main streets. In some cities, small markets also developed around the gates, these being favorable sites for periodic market fairs serving people from the rural areas. Cultivated lands, which still exist in many walled cities of western China, were found mostly in the area close to the wall. Due mainly to this dispersed distribution of shops and markets, the concentric zoning pattern of land-use inside the city was less evident than it is in many urban centers of the Western world today.

Although the growth of hsien capitals influenced their general structure, it left the central part of the town almost untouched. The center retained its distinct form and character, and remained an organic part of the whole plan even when new rectangular units were added to the existing pattern of squares and rectangles. The extension pattern of walled cities varied from one to another according to their population growth and regional character, but in most cases it followed generally four types: (1) In lowland areas, where a modern transportation system had not been developed, the extension was often carried out systematically along country roads radiating from the four gates (Fig. 12). (2) In areas of rugged relief, the suburbs surrounding old districts tended to follow the limited level terrain and the main highways, the new settlements often being shaped by contour lines of the local relief (Fig. 13).

(3) If the original town stood a few miles away from a river, the new settlements often extended all the way down to the bank, so as to have better access to the waterway. (4) Since the latter part of the nineteenth century, many hsien capitals have been linked with railroads. Railroads in most cases have not penetrated into the walled cities, perhaps partly because of the objections of the peasants and the local gentry who were concerned with the observance of local "feng shui." Instead, stations were usually about one or two miles outside city walls, and gradually served as the nuclei of new business districts. Thus, many hsien capitals grew into the multiple-nuclei type of city arrangement.[34]

THE FUNCTION OF THE HSIEN CAPITAL

As mentioned before, the hsien capital was originally established primarily for local administration. For more than 2,500 years since the Chou dynasty, however, the hsien capital has gradually evolved into a regional center, economically and culturally, as well as politically.

The hsien itself was a region, not so much in the sense that its boundary coincided with physical barriers, but rather in the sense that its social, economic, and cultural phenomena were similar, even though many hsien did happen to coincide closely with an intermontane basin or a flood plain. E. G. R. Taylor defines a region as "an area which is homogeneous in respect of the group consciousness, organization and interdependence of its people and is distinct from surrounding areas."[35] It is in this sense that a hsien was functionally a geographic region. Chinese society has been called "family-linked" and "earth-bound," but it was also "regional-minded." If the Chinese empire were considered to be an organic body, the hsien unit might be conceived of as an individual cell, growing around an active nucleus, the hsien capital.

The fact that the hsien was a region can be best illustrated by its group feeling. In many communities of South China, especially along the coastal provinces of the southeast, the hsien was a linguistic unit. Differences in accents and dialects from hsien to hsien were partly responsible for group consciousness. In the eastern part of Chekiang province, a traveler from one hsien, with its particular dialect, would often be considered a foreigner by the people of other hsien. Almost all of the Chinese in Chinatowns of the United States were originally from Taisan hsien in the Province of Kwangtung. They spoke the dialect of Taisan which was not easily understood by a person from Canton, though Canton is only 40 miles from Taisan. Likewise, each of the many Chinese communities in various countries of southeast Asia was settled mainly from a particular hsien.[36]

The defining of regions based on hsien was also expressed in the enormous compilation of Hsien Chih (the local histories of a hsien), which was probably the most important common task of all the hsien throughout the country and was undertaken every few decades. The Hsien Chih were written and edited by the most prominent scholars in each hsien and contain a great treasury of local records which were significant enough to show the local spirit and regional pride of the hsien in detail. The people of a certain hsien were often so proud of a native scholar or a statesman who had won nationwide fame that they let him bear the name of the hsien. Therefore, K'ang Yu-wei, a political reformer, was also known as K'ang Nan-hai, since he was born at Nanhai hsien of Kwangtung province. Li Hung-chang, one of the few talented diplomats of the late Ch'ing dynasty, was often called Li Ho-wei, Howei being his native town in Anwhei province.

Unlike many county seats in the United States, which serve only as county administrative centers and are not necessarily economic or cultural centers, hsien capitals had multiple functions for the regions which surrounded them. These functions can be explained in the following terms:

1. *Administrative Headquarters.* The ad-

[34] Harris and Ullman have classified cities into three types: concentric zone type, sector type, and multiple nuclei type, according to their internal structure. See Chauncy D. Harris and Edward L. Ullman, "The Nature of Cities," *Annals*, American Academy of Political and Social Science, Vol. 242 (November, 1945).

[35] Robert E. Dickinson, *City Region and Regionalism* (London: 1947), p. 9.

[36] Ta Chen, "Chinese Migration, with Special Reference to Labor Conditions," *Bulletin of the United States Bureau of Labor*, No. 340 (1923), p. 27.

ministrative character of the hsien capital underwent little change through 2,500 years. To the rural people, the city wall symbolized authority and dignity, since the magistrate inside the wall represented the emperor and the imperial government. The multiple duties borne by the magistrate could be traced back to the Han dynasty.

> All kinds of hsien heads contribute to governing the people. They display goodness and encourage righteousness; they prevent evil and punish the wicked; they settle litigation and equalize taxes; and they succor the people in their seasonal occupations. In autumn or winter they collect the revenue and send up accounts to the commandery or kingdom to which they are subordinate.[37]

Thus, such an official was not only a fiscal officer, charged with collecting revenues and taxes from the land and other sources. He was also the justice of the peace, the mayor of the municipality, the hsien sheriff, the educational head, the agricultural adviser, the census conductor, and the chief engineer for irrigation projects and water control. With the aid of garrison troops, which were usually stationed in the walled city, he was also the chief in command of all local military campaigns against bandits or rebels. But these modern titles do not do justice to his almost unlimited activities, as he was expected to become acquainted with every phase of social, economic, and political life in the hsien. His local duties were more extensive than those of the Home Secretary of England.[38] The family was the basic unit of Chinese society, and the magistrate represented a paternal government with its subject families comprising a big family, the hsien. Therefore, the maxim of the magistrate was "love your people as your sons," and his popular name was "fu-mu-kuan" (father and mother official). On the other hand, as the children of a family had to practice filial piety towards their parents, the subjects of the magistrate were expected to have obedience and loyalty to their master.

The hsien capital was often located at the geographical center of the hsien; according to traditional Chinese ideology, this location was intended to reflect its supreme authority over the entire hsien territory. In the same way China's supposed centrality within the known world, as the "middle kingdom," was thought to reflect or even to be one cause of the cultural and political superiority of the empire. The hsien capital as regional center has been repeatedly illustrated by Chinese cartographers for many dynasties. Figure 14 is a pictorial relief map of Shien Han hsien in Szechuan province made during the early nineteenth century.[39] In this map, all of the relief features have a centripetal arrangement so that the perspective view is from the centrally located walled city. City dwellers considered the walled city as the center of the hsien not only in a geographical sense but also in a psychological sense.

2. *Cultural Center.* Since the gentry class, the only educated group in traditional Chinese society, lived in the hsien capital, the hsien capital became the cultural center for its area almost from the time of its establishment. Traditional Chinese education consisted mainly of the study of classics in the frame of the Confucian ideology, and every hsien capital had a Confucian temple which symbolized higher learning and ethical morality. Before the modern educational system was adopted, the Chinese student learned classics mainly at home through a tutor who also instructed students from neighboring families. This tutoring system was found in the city and in the rural area. The Shu Yuan system, in which groups of students studied under learned scholars, was found mainly in hsien capitals. This system has probably been the most important educational institution for preserving the traditional cultural pattern, ethical ideology, classical literature, and ancient philosophy for many centuries. Many well-known scholars, philosophers, and statesmen through history have been instructors of Shu Yuan in their native hsien capitals. Shu Yuan did not give any academic degrees, but did help to prepare students for the imperial examinations. The imperial government of China, in order to choose the most learned men for civil service, used the traditional civil examination system which dated probably from the Han dynasty. The system consisted of a ladder of examinations, starting at the

[37] Ssu-ma Piao, *Hou Han Chih*, Vol. 28, p. 7.

[38] Leong and Tao, *Village and Town Life in China* (London, 1914), p. 35.

[39] The map is a photographic print of the original copy which is attached to Volume II of *Shien Han Hsien Chih* (1815).

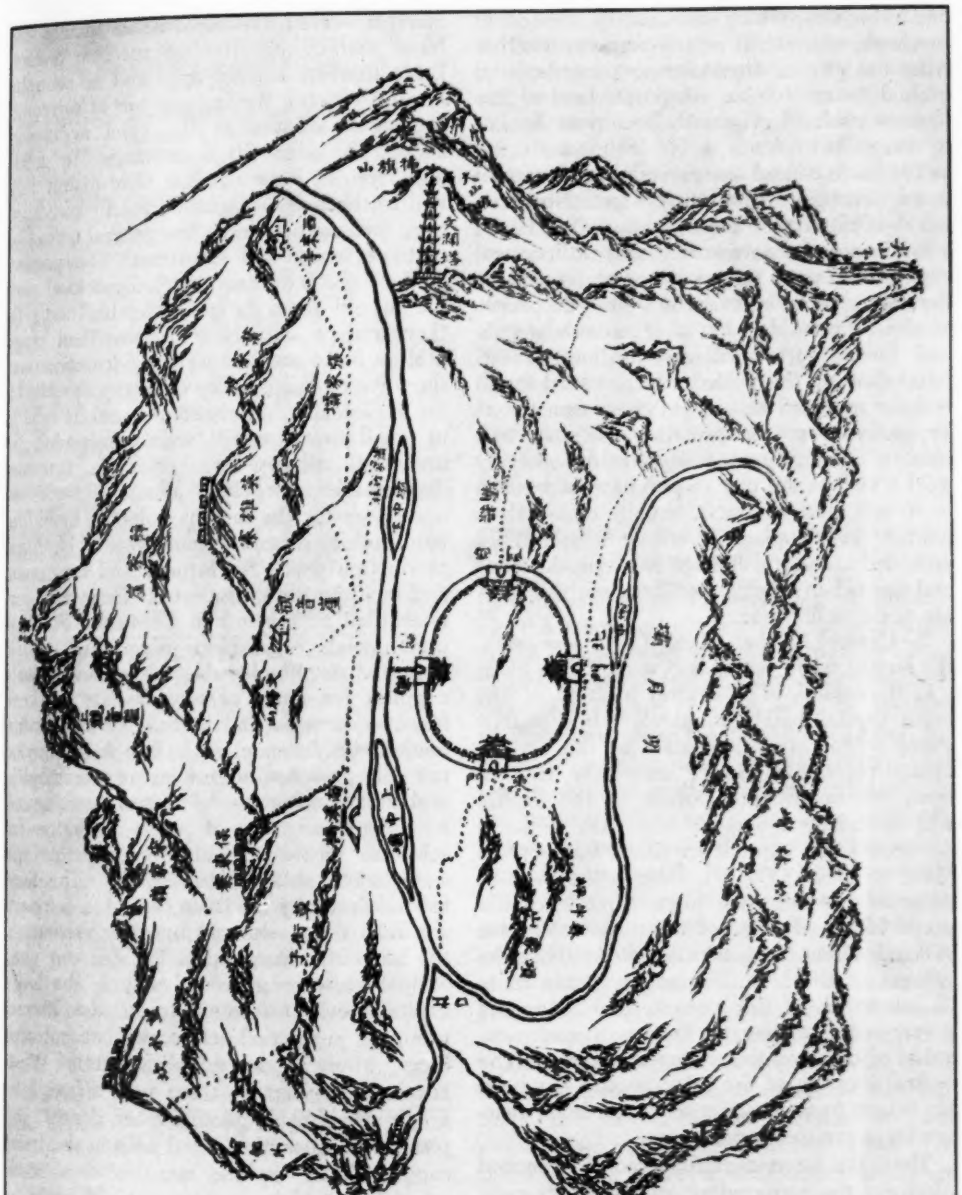


FIG. 14. The hsien as viewed perspectively from a centrally located walled city. (Source: *Shien Han Hsien Chih*, 1815.)

hsien level and culminating with the nation-wide imperial examination. The barracks-like buildings of the examination halls remain today in many hsien capitals of China. After the modern school system was adopted (about 1910), almost every hsien established a sec-

ondary school, which was usually located in the hsien capital. Thus any student, whether from the city or from the countryside, who wished to receive an education beyond the elementary level eventually became a dweller in the walled city.

The hsien capital also served as a recreation center for the surrounding area. Near the business district of the city was often found a local opera house where stories of historical significance were staged as entertainment for the gentry group and for the well-to-do people of the countryside. For the common people and farmers from villages, teahouses scattered through the walled city provided facilities for rest and relaxation; they also served as centers for the spreading of news and rumors. The country folk, who occasionally paid a visit to the city to purchase necessities or to sell farm products, usually ended their journey in a teahouse, where a storyteller provided a cultural linkage between the rural and the urban worlds, and between the literate and the illiterate.

3. *Central Market.* Except in a few cases, the largest trade center and market of a hsien was its capital. The central location of the hsien capital had perhaps the same significance as the central location of the departmental capitals of France, where the distance from the department border to the capital was ordinarily a one-day trip. The distance between hsien capitals in China varies from place to place (Fig. 3), though the average distance between two hsien capitals of the north China plain or of the middle Yangtze valley is about 50 to 60 kilometers. Thus the villages on the hsien border are within 25 to 30 kilometers of the hsien capital. A strong porter with an average burden is easily capable of covering this distance in a day. The central location of the hsien capital has been important in its position as an all-hsien market for a pre-industrial society.

Thus, the hsien capital served as a central place for the surrounding areas, and its position was approximately equivalent to the *Kreisstadt* in Christaller's hierarchy.[40] Below the level of the hsien capital, two kinds of markets served the rural area, namely, the basic market and the intermediate market. Basic markets usually appeared in centrally located villages, serving a group of surrounding small villages as the chief organizing agents for inter-village exchange. In goods and services, basic markets were mainly limited to the bare necessities of daily consumption, such as food and a few general items like matches, soap, and cigarettes. The paucity of their goods and services led the local people to call them "green vegetable" markets, though they actually sold more than vegetables. Basic markets were service centers of the lowest rank, roughly equivalent to *Marktort* in Germany. Intermediate markets existed in small towns which were surrounded by numerous villages, large or small. Intermediate markets provided almost all types of necessities for the local population, including food, sundry supplies, general wares, clothing, production goods for farmers and craftsmen, and facilities for amusements. Their rank corresponded to *Amtsort* in Germany. For luxuries, goods of a more permanent nature, imported manufactured goods, and wholesale supplies for basic or intermediate markets, however, people had to turn to the central market, the hsien capital. The hsien capital not only provided more items of merchandise and a larger number of stores, but also afforded a wider range of quality and prices for selection as well as supplying articles and services not obtainable elsewhere. The services rendered by the hsien capital in comparison with those rendered by other markets in the hsien are illustrated in Table 1.

Book stores were found only in the hsien capital, since the majority of the literate group, as mentioned before, was concentrated there. Many service establishments of Western influence, such as the post office, telegraph office, bank, photographer, dentist, and pharmacy, were also found only in the hsien capital. Thus, on one occasion or another, people within the service area of basic or intermediate markets had to patronize the hsien capital. The frequency of such visits depended on the needs of the people and the accessibility of the city. The central market of the hsien capital usually had two service areas: a primary area, including nearby villages where people attended market regularly

[40] Walther Christaller, *Die Zentralen Orte in Süddeutschland* (Jena: Gustav Fischer, 1933). Translated at Bureau of Population and Urban Research, University of Virginia, by C. Baskin (1954).

TABLE 1. NUMBER OF SERVICE ESTABLISHMENTS IN THE HSIEN CAPITAL AND OTHER MARKET CENTERS OF TUNG KUAN HSIEN IN SHENSI PROVINCE (1944)[1]

Type of service	In Hsien capitals	In other market centers						Total
		A	B	C	D	E	F	
General store	56	13	15	10	9	6	5	114
Grain store	3	0	1	2	2	1	0	9
Herb drug store	12	1	1	5	6	5	1	31
Restaurant	7	6	5	5	3	3	2	31
Inn	6	2	2	3	0	0	0	13
Dyeing shop	2	0	0	2	1	1	0	6
Doctor	3	1	0	1	0	0	0	5
Tailor	4	3	0	3	3	1	0	14
Leather shop	5	0	1	2	1	0	9	18
Jewelry shop	2	0	0	1	1	0	0	4
Public bath	1	1	0	1	0	0	0	3
Barber	3	2	1	2	3	1	1	13
Books & stationery	5	0	0	0	0	0	0	5
Bank	2	0	0	0	0	0	0	2
Photographer	4	0	0	0	0	0	0	4
Post office	1	0	0	0	0	0	0	1
Telegraph office	1	0	0	0	0	0	0	1
Pharmacy	2	0	0	0	0	0	0	2
Dentist	3	0	0	0	0	0	0	3
Total	122	29	26	37	29	18	18	279

[1] Compiled from description in *T'ung Kuan Hsien Chih*, Vol. XII, "Industry and Commerce" (1944).

or at least frequently, and a secondary area encompassing more distant villages where people came to market only occasionally for items scarce in their own neighborhood markets. Therefore the hsien capital covered a much greater service area than did the basic market or intermediate market. Ching-kun Yang made a field study concerning the size and spacing of rural service centers in Shantung province.[41] His findings testify to the important position of the hsien capital in the central-place hierarchy of China as follows:

		Service area	
Central place	*Number*	*Mean radius*	*Average population*
Basic market	10	1.3 miles	3,847
Intermediate market	18	2.5 miles	8,513
Central market (hsien capital)	1	8.2 miles	48,454

Access from rural areas to the hsien capital has been improved during the last few decades. Besides water transportation, the hsien capital has often been connected with the rural areas by highways and hsien roads (Fig. 15), especially the latter, which were designed by the local government to radiate from the city to the countryside in all directions, even though most of them were not wide enough for automobiles. When the railroads came to China, they were often arranged to serve as modern transportation lines connecting hsien capitals. When a railroad passed through the territory of a hsien, the only cargo and passenger station was often at the capital. As a result, country folk from all over the hsien who wished to take trips to other provinces by rail had to come to the capital to take their trains. Therefore, the hsien capital functioned in many ways as the only window through which the rural communities of China could communicate with the outside world.

[41] Ching-kun Yang, *A North China Local Market Economy* (New York: Institute of Pacific Relation, 1944).

EFFECT OF HSIEN CAPITAL ON URBANIZATION OF CHINA

As mentioned at the beginning of this article, the hsien capital played an important role in the development of urbanization in China before the middle of the nineteenth century. Since the opening of treaty ports by Western powers, the urban hierarchy of China has changed greatly. The imperial capital has lost its position as the most populous city in the nation while port cities have gradually found themselves among the largest urban centers in the world. Nevertheless, the walled cities along the main waterways and along the coast have provided sites for many treaty ports which have become metropolitan in character. Shanghai, Tientsin, Wuhan, Chungking, Ningpo, Amoy, and Chefoo, among other ports, were all originally hsien capitals.

Since the Communist regime was established in 1949 and rapid economic development has begun, China appears to have entered into a new stage of urbanization. The hsien capital, however, still serves as a base for the growth of urban population. Among 189 communities announced as "cities" by Peking up to 1958,[42] 160 were originally hsien

[42] Theodore Shabad compiled a list of Chinese urban places and their populations up to 1958 from Russian sources. See Theodore Shabad, "The Population of China's Cities," *Geographical Review*, Vol. XLIX, No. 1 (January, 1959), pp. 32–42.

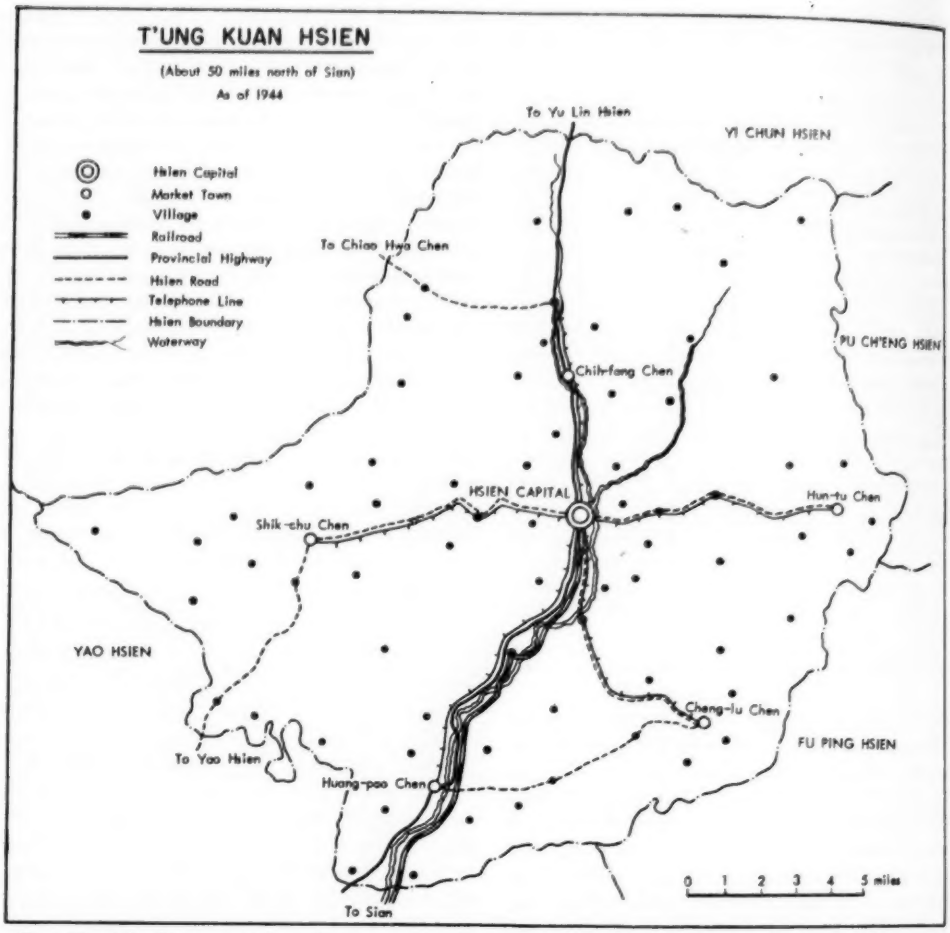


FIG. 15. The hsien capital as the communication center. (Source: *T'ung Kuan Hsien Chih*, 1944.)

capitals.[43] During the last ten years, as a result of the new regime's policy, which has

[43] Comparing the list of cities by Shabad with the list of hsien capitals in *Chung Hua Min Kuo Hsing Cheng Ch'u Yu Chien Piao* ("An Outline of Administrative Units of the Republic of China"), compiled by the Interior Department of China (Shanghai, 1947), leaves 29 cities which are not hsien capitals in origin, as follows:

Anwhei	Pengpu
	Tungkwanshan
	T'unch'i
	Maanshan
Honan	Chumatien
	Loho
	Chiaotso
	Hopi
	Pingtingshan
	Sanmensia
Inner Mongolia	Baynahot
	Hailar
	Ulanhot
Kansu	Paiyin
Kiangsi	Fuchow
Kwangtung	Pakhoi
	Swatow
Manchuria	Dairen
	Ichun
	Hokong
	Kisi
	Kungchuling
	Port Arthur
	Shwangyashan
Shantung	Ts'ingtao
	Weihai
Sinkiang	Karamai
Tibet	Lhasa
Yunnan	Siakwan

put emphasis on the economic development of the northwestern territory, a great many urban centers have been developed there and new economic activities have been added to the cities in the area; for example, Paotow has become a new iron and steel center; Paochi, a railroad center; Lanchow, a cultural and industrial center; and Yumen, a center of petroleum refineries. These new urban settlements were all hsien capitals in origin and had great military and trade significance for many centuries, especially in the Han and Tang dynasties. Therefore the growth of these cities, in a historical sense, is a revival of the old walled cities along the ancient Silk Road and caravan routes. In view of the new regime's rapid railroad construction, its establishment of steel plants on a regional basis, its extensive survey projects by geological teams, and its ambitious aims for over-all industrial development throughout the nation, it is likely that China will experience rapid urbanization during the next few decades. Many hsien capitals, with their carefully selected sites, their commanding positions in the transportation networks, their historical background, their existing buildings and facilities, and above all, their widespread distribution throughout almost all of the areas which are suitable for any development of a concentrated economy,[44] are likely to grow and to become among the largest urban centers in the nation.

[44] Urbanization, in a spatial sense, is a concentration of people, buildings, and machines within a relatively small area. See Richard U. Ratcliff, "The Dynamics of Efficiency in the Locational Distribution of Urban Activities," in *The Metropolis in Modern Life*, ed. by Robert Moore Fisher (New York, 1955), pp. 125–48.

MISSIONS AND MISSION LANDS OF ALTA CALIFORNIA

R. LOUIS GENTILCORE
McMaster University

THE Franciscan missions provide the best-known and most obvious sign of California's past. They are both a distinctive landscape feature and a link with the events that transformed that landscape. In the fifty-four years from the founding of the first to the last of the missions, there had been transferred to the coastal bays and valleys of California the significant domesticated plants and animals of Spain. Horticulture and gardening were introduced and confined almost exclusively to the missions. From the Mexican border to the valleys north of San Francisco, wheat became the dominant crop. Over this whole territory and well beyond all cropland ranged thousands of cattle and sheep. Trade was an important activity and set the stage for the large-scale movement of hides from California's range lands. Material prosperity was evident, reflected not only in imposing buildings and workshops but also in large, active Indian populations, devoted to new occupations, such as animal herding, field-crop and fruit production, stone-masonry, carpentry, tanning, and weaving. In the words of Peattie, the missions were "like a piece of Europe set down amidst the chaparral."[1]

The transformation of landscape began with the selection of mission sites and the erection of the first buildings. Accordingly, this study considers first some of the conditions, physical and cultural, under which the missions were established. Mission sites and locations are related to topography, soils, water supply, and population distribution. With establishment, a new economy appeared. The rest of the paper is concerned with this economy. Agriculture, ranching, trade, and horticulture are described. They are related to environmental conditions and compared from one mission to another. The progress made by the missions is well documented in reports compiled by the missionaries. The wealth of material available in these reports would be sufficient reason in itself for undertaking a description of the geography of the missions.

[1] R. Peattie, *The Pacific Coast Ranges* (New York, 1946), p. 4.

ESTABLISHMENT OF THE MISSIONS

Sites and Distribution

The founding of a mission and its subsequent development followed a pattern born of years of experience, accumulated throughout the whole of Spanish America. The missions were pioneer institutions.[2] Their establishment involved exploration of territory, pacification of the native inhabitants, and the introduction of a new civilization. Wherever a frontier needed to be crossed and new land occupied, missions led the way. Thus, Jesuit missions opened the frontier lands of northwestern Mexico. Settlement moved northward and eventually, from the well-established missions of Baja California, the occupation of Alta California took place.

In Alta California mission activity was carried on by the Franciscans. Their accomplishments, remarkable in many ways, reflect not only their own abilities but the effectiveness of the procedures they followed in establishing the missions. For example, a good deal of attention was paid to the proper location of each mission. Royal instructions explicitly directed the selection of sites. They specified that each mission should have ample land for agriculture, access to water supply, and Indian populations near by. Other items mentioned were availability of timber and access to the ocean for trade. In short, the selection was based on appraisal of topography, resources, and population.

The skill with which the appraisal was made is evident from a comparison of the location of the missions and the present distribution of population (Fig. 1). The mission chain coincides with the major occupied areas of coastal California which contain the bulk of the state's population. The Central Valley was scarcely touched, but this was not an oversight. Plans were formulated and preliminary attempts were made to penetrate the heart of the Tulare country. Here were large

[2] H. E. Bolton, "The Mission as a Frontier Institution in the Spanish-American Colonies," *American Historical Review*, Vol. XXIII (1918), pp. 42–61.

FIG. 1. Distribution of missions with relation to major populated areas in California today. The mission chain, numbered from 1 to 21 south to north, coincides with the occupied areas along the coast. The Central Valley was not penetrated although it was included in the settlement plans of the Franciscans. Areas of occupation are taken from Burcham.

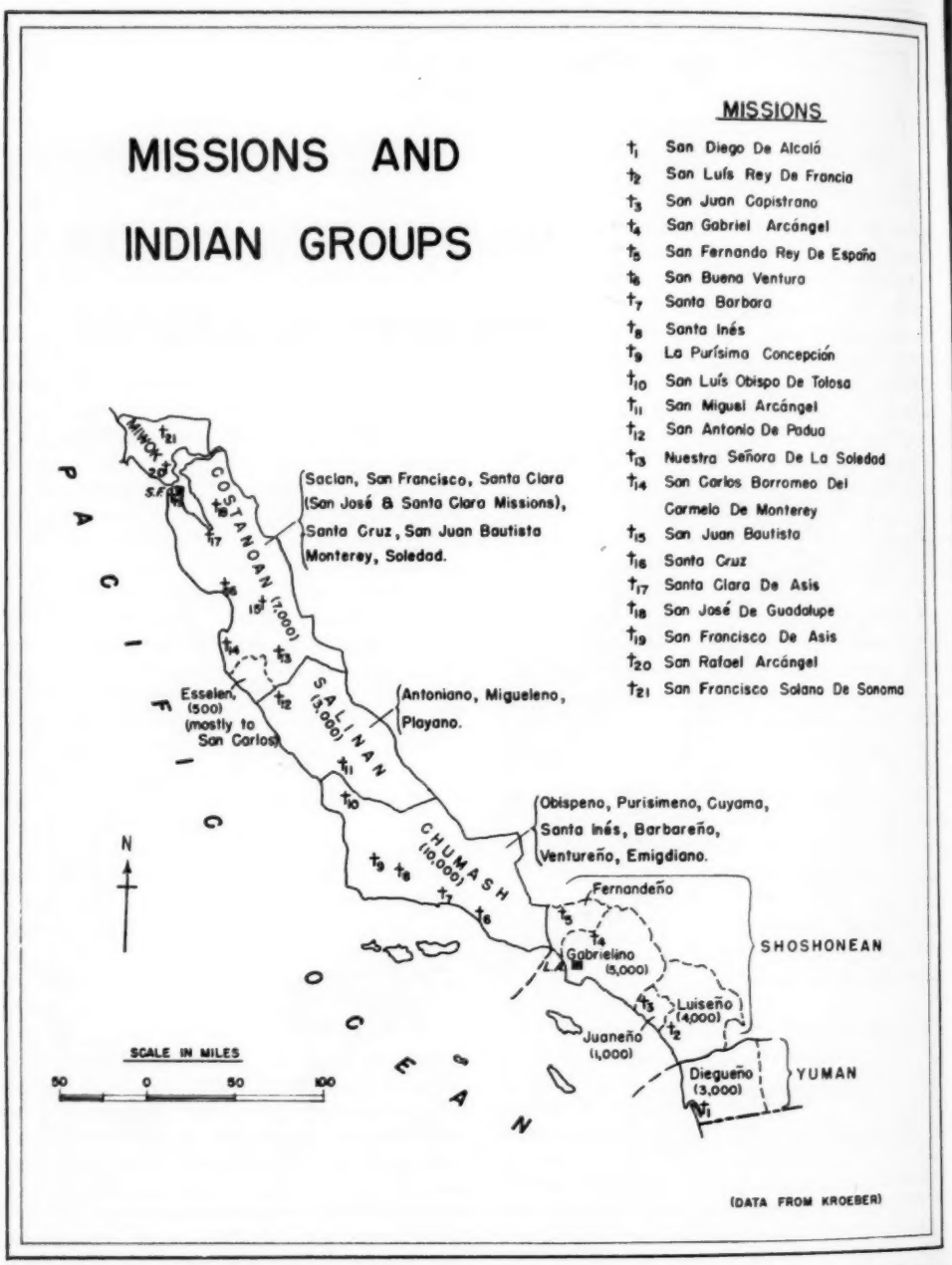

MISSIONS AND INDIAN GROUPS
MISSIONS
†1 San Diego De Alcalá
†2 San Luís Rey De Francia
†3 San Juan Capistrano
†4 San Gabriel Arcángel
†5 San Fernando Rey De España
†6 San Buena Ventura
†7 Santa Barbara
†8 Santa Inés
†9 La Purísima Concepción
†10 San Luís Obispo De Tolosa
†11 San Miguel Arcángel
†12 San Antonio De Padua
†13 Nuestra Señora De La Soledad
†14 San Carlos Borromeo Del Carmelo De Monterey
†15 San Juan Bautista
†16 Santa Cruz
†17 Santa Clara De Asis
†18 San José De Guadalupe
†19 San Francisco De Asis
†20 San Rafael Arcángel
†21 San Francisco Solano De Sonoma
MIWOK
COSTANOAN (7,000)
Sacian, San Francisco, Santa Clara (San José & Santa Clara Missions), Santa Cruz, San Juan Bautista Monterey, Soledad.
Esselen (500) (mostly to San Carlos)
SALINAN (3,000)
Antoniano, Migueleno, Playano.
CHUMASH (10,000)
Obispeno, Purisimeno, Cuyama, Santa Inés, Barbareño, Ventureño, Emigdiano.
Fernandeño
Gabrielino (5,000)
SHOSHONEAN
Luiseño (4,000)
Juaneño (1,000)
Diegueño (3,000)
YUMAN
PACIFIC OCEAN
N
SCALE IN MILES
50 0 50 100
(DATA FROM KROEBER)

Fig. 2. Distribution of missions with relation to native regions. One of the factors in the establishment of a mission was the presence of fairly large Indian populations. Tribes were named, in turn, after missions established in their territories. Numbers in parentheses indicate Kroeber's estimates for the various populations at the time of the coming of the Spaniards.

Indian populations which not only threatened the coastal missions with attack but, more important, offered prospects of an abundant harvest of souls to the eager missionaries. Subsequent events were to deny these plans, and the second chain of missions never came into being. But the first chain was well established, and its close relationship to Indian groups along the coast is shown in Figure 2. Reference will be made later to some of these groups.

Topography

All of the missions were located among the ranges of California's coast: the Coast, Transverse, and Peninsular ranges (Fig. 3). Among the sites selected were valleys, alluvial fans, and coastal plains, located some distance from marshy flats and in most instances close to the sea. In some cases the position was to become a focus for important interior routes. San Gabriel, for example, was midway between San Diego and Santa Barbara and commanded mountain passes behind it. San José similarly guarded the natural highway entering the San Joaquin Valley from the coast.

Three missions were located within the Peninsular Ranges—San Diego, San Luís Rey, and San Juan Capistrano (Fig. 3). All were close to the sea (three to six miles) and occupied valley bottoms and slopes along streams whose channels had been cut from marine terraces. The most impressive of these valleys is that of the San Diego River, 300 feet deep and one-half to three-quarters of a mile wide. Mission San Diego was at an elevation of 225 feet, about 25 feet above the stream, and well above the delta of the river, six miles away. In contrast, San Luís Rey was lower and, more significantly, had a wider valley and larger areas of level land for agriculture. The most restricted of the valley sites was that of San Juan Capistrano. Here the trench valleys of two creeks coalesce. The valleys are narrow, and room for expansion was limited.

The Transverse Ranges contained the six missions from San Gabriel to La Purísima Concepción, none of which was more than 25 miles from the ocean. Three were located on alluvial fans: San Gabriel and San Fernando on fans below the San Gabriel Mountains and Santa Barbara below the Santa Ynez range. The latter had the most cramped location with a small amount of level land between the mountains and the sea. San Gabriel and San Fernando had no such limitations, and San Gabriel in fact was able to expand its operations over large areas in the so-called San Gabriel Valley. The advantages of the site were recognized very early by the founder of the mission himself, Father Junípero Serra, who described its facilities as "beyond dispute the most excellent of all discovered."[3] Another early description by the inveterate chronicler Father Paloú cited the situation of the mission "on the slope of a hill in [a] valley . . . which is very spacious with plenty of land."[4]

The other three missions had valley locations. La Purísima was actually on the lower parts of an alluvial fan, but its occupied area and the first mission site were below this, on the flat lands of the upper Santa Ynez River. Downstream, Mission Santa Inés lay in the structural valley of the river, between the San Rafael Mountains to the north and the Santa Ynez Mountains to the south. Back along the coast, east of Santa Barbara, was one of the few locations on a well-developed coastal plain, that of Mission San Buenaventura (Ventura). It occupied the lowland at the mouth of the Ventura River at an elevation below fifty feet. The river flows from the Santa Ynez Mountains, twelve miles away, in a narrow flood plain whose lower parts are deeply filled with alluvium. To many of the mission observers this was "the fruitful valley," and Vancouver noted its "reputation of being amongst the most fertile of the establishments in the country."[5]

The remaining missions occupied interior and coastal sites in the valleys of the Coast Ranges. The largest intermontane valley, the Salinas, had two missions on it, San Miguel and Nuestra Señora de la Soledad. The former was in the narrow upper part of the valley where the floor is only one-half mile wide. Below the junction of San Antonio Creek, the Salinas Valley widens and Soledad was in this

---

[3] Z. Engelhardt, *San Gabriel Mission* (San Gabriel, 1927), p. 10.

[4] Francisco Paloú, *Historical Memoirs of New California*, trans. and ed. by H. E. Bolton (4 vols.; Berkeley, Calif., 1926), Vol. III, p. 217.

[5] George Vancouver, *Vancouver in California 1792–94*, ed. and annotated by M. E. Wilbur (3 vols.; Los Angeles, 1953, 1954), Vol. II, p. 228. This material was first published in Vancouver, *Voyage of Discovery to the North Pacific Ocean and Round the World . . . 1790–95* (3 vols.; London, 1798).

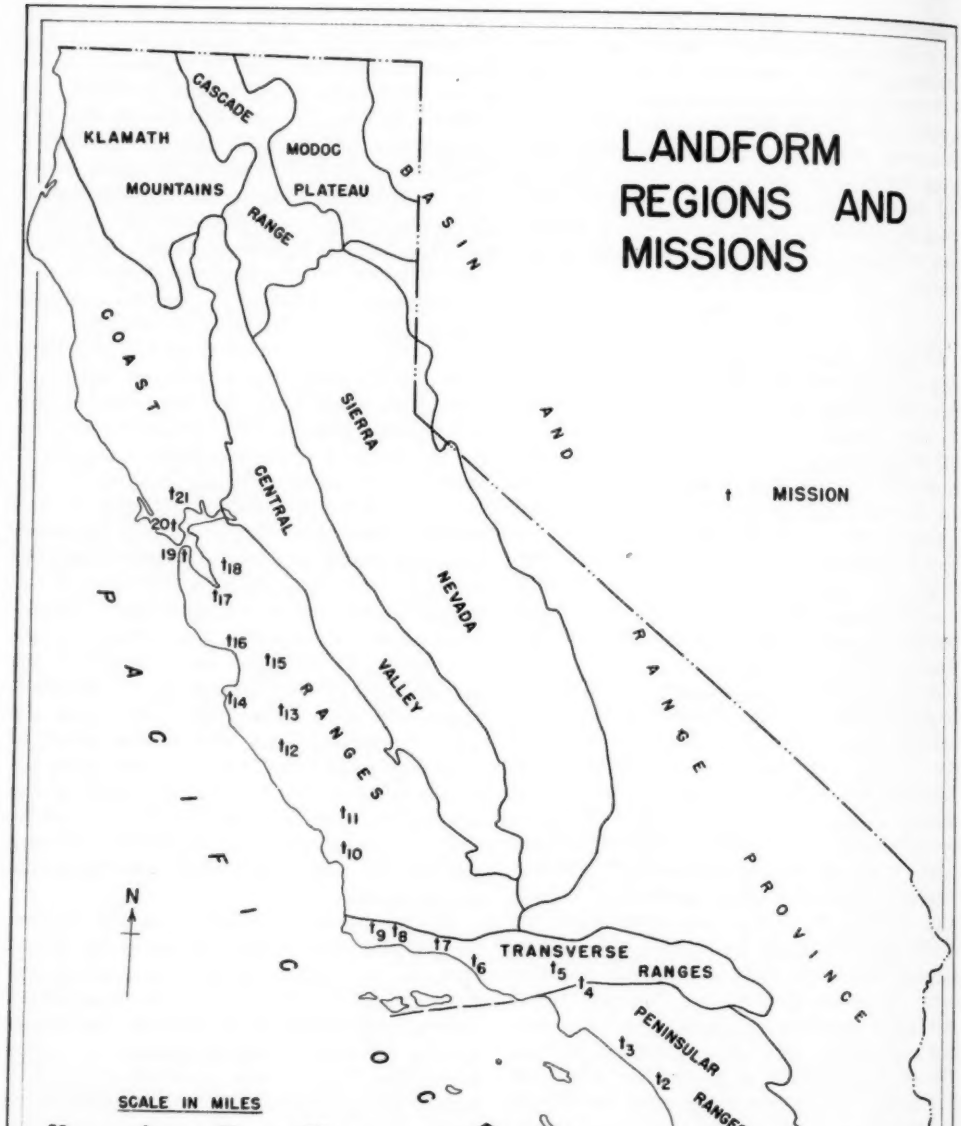


Fig. 3. Distribution of missions with relation to the landform regions of the state. All of the missions were located among the ranges of California's coast—the Coast, Transverse, and Peninsular ranges. Sites selected were valleys, alluvial fans, and coastal plains. No location was very far from the sea.

section where the floor is about five miles wide. A third mission, San Antonio, was on the San Antonio tributary of the Salinas but well within the interior at an elevation of over 1,000 feet and effectively separated from the main valley by mountains. The site is still isolated and the beauty of the setting, in a small valley surrounded by high mountains, continues to captivate its visitors. Although the San Antonio Valley is not wide, there is easy access from it into other valleys nearby and the mission did not lack for level land. In the same way Mission San Luís Obispo, in the narrow upper parts of San Luis Obispo Canyon, now fully occupied by the city of San Luis Obispo, had below it the more extensive flat lands of the San Luis Valley near the coast. These, however, are low-lying and poorly drained and did not present the same advantages for occupancy as the San Antonio environs.

Another interior location was that of San Juan Bautista, along San Juan Creek, forming part of the large lowland area now known as the San Benito Valley. Three valleys meet here—those of the San Benito, Pajaro and Santa Clara rivers—in a plain that is two miles wide. Offering even better topographical facilities were the sites of Santa Clara and San José. Santa Clara, like Ventura, was on a large alluvial plain along the coast, in this instance a deltaic fan. Much of the area, however, could not be used because of marshy conditions. San José, on the lower slopes of Mission Peak, stood above the coastal marshlands and utilized valleys in the interior. Two of these were the present Livermore and Sunol Valley.

The remaining missions from San Carlos to San Rafael were on the coast, on slight elevations overlooking bays. None was well supplied with level land. San Carlos at the mouth of the Carmel Valley had only the limited area of its own valley. The same was true of Santa Cruz and its river San Lorenzo which contains small areas of bottomland. References to the small areas at San Francisco and San Rafael are common. The northernmost mission, San Francisco Solano, on the low parts of an alluvial fan above Sonoma Creek, was intended as a replacement for San Francisco and provided the valley lands not available at the latter mission. Duhaut-Cilly described the site "in the middle of a plain of great extent . . . everywhere watered" and went on to declare that "there are few happier sites and this mission might become a very important one in a future time."[6] His hopes were not fulfilled, but not because of topographic deficiencies.

Soils

On Figure 4 the most important categories of California soils—well drained basins, valley floors, and piedmont areas—are outlined for the coastal parts of the state. The comparison of this distribution with that of the missions indicates how astute the Franciscans were in their judgment of soil. It was this selection that enabled Humboldt to write that "the soil of New California is as well watered and fertile as that of Old California is arid and stony."[7] Missions with extensive alluvial lands, such as San Gabriel, San Buenaventura, and Santa Clara, were especially well favored. Vancouver was very much impressed with Santa Clara. "It is situated," he wrote, "in an extensive fertile plain, the soil of which . . . is a rich black productive mould, superior to any I had before seen in America."[8]

In many places, although the soil was alluvium, it was limited in quantity. Reference has already been made to such locations. San Luís Obispo may be cited again. In 1787, Governor Fages could write that its soil was "of such good quality and so rich that it makes up for the insufficiency of rain."[9] The Governor had hopes that Obispo would keep pace with San Gabriel in production so that these two missions could provide grain to assist in the establishment of other missions, eliminating the need for imports from Mexico.[10] For a few years, these hopes were fulfilled but as the number of missions grew and

---

[6] A. Duhaut-Cilly, "Account of California in the Years 1827–28," *Quarterly, California Historical Society*, Vol. VIII (1929), p. 240. The "Account" is a translated excerpt from Duhaut-Cilly's "Voyage Autour de Monde, Principalement à la Californie . . . pendant les années 1826, 1827, 1828 et 1829" (Paris, 1834).

[7] A. Humboldt, *Political Essay on the Kingdom of New Spain* (4 vols.; London, 1811), Vol. II, p. 340.

[8] Vancouver, *op. cit.*, Vol. I, p. 36.

[9] Z. Engelhardt, *Mission San Luís Obispo* (Santa Barbara, 1933), p. 39.

[10] Pedro Fages, *Description of California . . . made for the Viceroy in the Year 1775*, trans. by H. I. Priestley (Berkeley, Calif., 1937), p. 43.

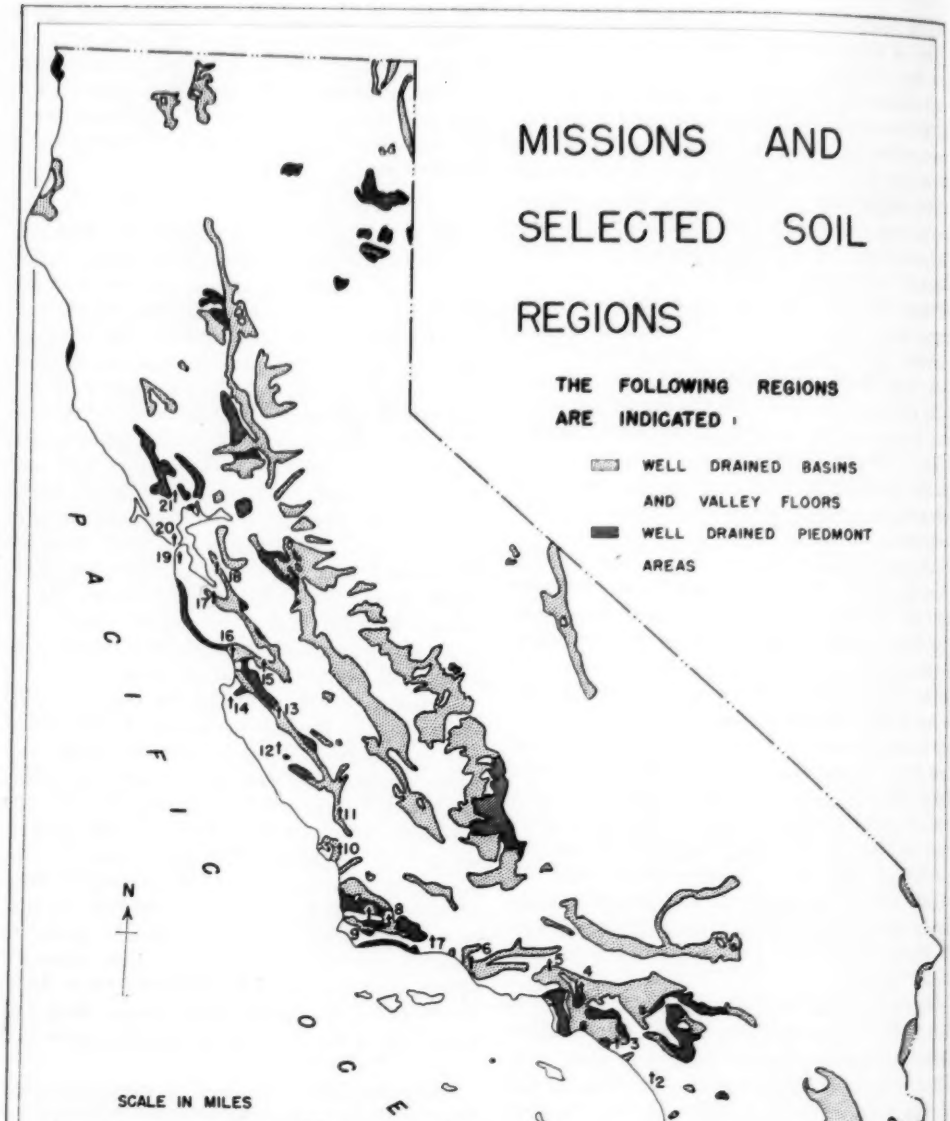

MISSIONS AND SELECTED SOIL REGIONS
THE FOLLOWING REGIONS ARE INDICATED :
WELL DRAINED BASINS AND VALLEY FLOORS
WELL DRAINED PIEDMONT AREAS
PACIFIC OCEAN
N
SCALE IN MILES
50 0 50 100
SOIL INFORMATION FROM STORIE AND WEIR

FIG. 4. Distribution of missions with relation to selected soil regions. The mission locations coincided with some of the best soils available in the coastal areas—the well-drained basins, valley floors, and piedmont areas. The selection of mission sites obviously involved astute evaluation of soil resources by the Franciscans. Soil regions from Durrenberger and Byron: *Patterns on the Land.*

as the lands of Obispo became fully utilized, other places became more important producers.

Even more common than reports on the fertility of the soil were those of its "sterility," which could be a reference either to soil or to drought conditions. "The territory is more sterile than any we have seen," wrote Father Crespí of the San Diego site,[11] and Father Paloú related how land better suited for cultivation was sought in the area to no avail. Specific features of "sterility" noted in early reports include alkalinity on some of the lands of San Juan Capistrano and San Luís Rey, sandy conditions around San Carlos, droughtiness at Soledad, and "the barrenness" around San Francisco. The latter is probably a reference to sandy soils. Vancouver pointed out, however, that the mission itself was in a pocket of soil in which sand and heavier materials were mixed, and that this was the most fertile spot on the peninsula.[12]

The utilization of soils over a long period of time naturally brought about changes not only in the soils but in the whole physical geography at the mission sites. A graphic description of such changes for Mission San Juan Capistrano is presented in a manuscript on Orange County. Describing the irrigated fields along the two major streams in the mission lands, the writers point out that these farm lands were much more extensive in the mission period. At that time the creek bottoms were almost level with the rest of the valley floor. During mission occupancy and since then, "destructive floods . . . scoured away thousands of acres of good land adjacent to the channels, leaving behind nothing but wide, deep gravel beds."[13] The destructive nature of the floods is attributed to the removal of vegetation on the upper part of the watershed by grazing and by fires, and on the lower parts for firewood and to make room for cultivation. Similar activities took place at the other missions. For example, Smith notes that while Mission San Antonio flourished, the best timber on Santa Lucia Peak, particularly sugar pine (*Pinus lambertiana*) and incense cedar (*Libocedrus decurrens*), was cut down and the two species were almost destroyed.[14] The floods in the wake of this destruction would again remove substantial acreages. Beyond this, it is difficult to say to what extent the processes of soil removal can be extended to the whole mission area. Specific evidence, such as that for San Juan Capistrano, has not been worked out. It is enough to say that the appearance and extent of the lands available to the missions were not the same as they are today.

Water Supply

The most important physical factor in the selection of a mission site was the availability of water. It mattered little that soil was fertile and level land abundant unless sufficient water could be provided. Even in the brief accounts of a particular mission, whether an official report or a visitor's impression, reference is usually made to the water supply or to water problems.

The concern with water in the selection of mission sites is illustrated by Font's remarks about the Santa Barbara coast before the founding of missions there. He reported that there was "no place to establish a good mission because of the small amount of water which the arroyos carry, for many dry up in the course of the year."[15] In the exploration for the site of San Miguel, Father Sitjar noted specific localities where grain could be grown without irrigation and others where arroyos with sufficient running water could be tapped.[16] Paloú saw possibilities in what were to become the mission grounds of Santa Clara because of the presence of a "good stream of water . . . from which plenty of ground could be irrigated."[17]

The chief complaint in the established missions was lack of water. Paloú noted that early San Diego "lacks water to irrigate the plentiful good lands that it has." Drought ruined many crops. Of one of the first, Paloú

[11] Z. Engelhardt, *San Diego Mission* (San Francisco, 1920), p. 15.

[12] Vancouver, *op. cit.*, Vol. I, p. 21.

[13] "Orange County, California—Water and Irrigation," Works Progress Administration Research Project 3105 (typewritten manuscript in public library, Santa Ana), p. 6.

[14] F. R. Smith, *The Mission of San Antonio de Padua* (Stanford, Calif., 1932), pp. 36–37.

[15] Pedro Font, *Diary of the Second Anza Expedition, 1775–1776*, trans. and ed. by H. E. Bolton (Berkeley, 1931), p. 256.

[16] Z. Engelhardt, *San Miguel Arcangel* (Santa Barbara, 1929), pp. 3–5.

[17] Paloú, *op. cit.*, Vol. IV, p. 160.

wrote, "the rains failed at the best time and the seed was lost."[18] At Santa Barbara, in line with Font's observations just noted, Governor Fages wrote that cultivation had to begin "in spite of the lack of sufficient water."[19] Even along streams as sizable as the Salinas, there were serious deficiencies. "This water," declared one report for Soledad, "seems only for the livestock because it is too low for irrigation and . . . runs dry in June or July."[20] At San José, above the marshlands adjacent to San Francisco Bay, the chief source of water was an arroyo, five leagues northwest of the mission, but even this was "reduced to a puddle" in summer and the adjacent territories became useless because they had "scarcely no water."[21] It is refreshing and quite exceptional to have a mission described in opposite terms. For example, Duhaut-Cilly writing of San Francisco Solano described its lands as "everywhere watered and crossed by streams of fresh water."[22] Similar scattered phrases are used for some of the other missions but these do not refer to conditions over a long period of time. They are submerged under more persistent and more penetrating evaluations of water facilities.

The seasonal flooding of streams presented another water problem. The nature of the California climate, with its seasonal rains and the inability of streams and canyons to carry large volumes of water, made this inevitable. Conditions were aggravated as vegetation cover was removed. A graphic description of the kind of flooding that affected the missions is contained in a report to Governor Echeandía on the state of San Juan Capistrano. "The floods from the arroyos," it related, "are very destructive, for when there is snow in the mountains, the sun melts it, and the rains increase the flow; then the arroyos become rivers, which leave their bed, form other currents, and in receding leave nothing but stones and sand in the land, as the Indians, the missionaries, and all passers-by observe with tears in their eyes and pain in their hearts."[23]

Floods or lack of water forced the missionaries to relocate some of their missions. Six were moved at least once. Because of difficulties with floods and drought, San Diego was moved from the Presidio Hill up the San Diego Valley. In the transfer, directions were laid down by the Viceroy for increasing the efficiency of collecting water, at first by "small ditches from ponds of water or wells in the bed of a stream" and ultimately by "a firm dam from which (water) may be drawn for the future."[24] San Gabriel was moved from its original site on a bluff overlooking the Rio Hondo because of the destruction of floods. Its new site, eight miles upstream, was higher and drier. La Purísima was moved after the earthquake of 1812 to the northern side of the Santa Cruz River where it had access to better supplies of surface and underground water.

San Antonio was "compelled to move because the river ran dry." Its new location was about four miles upstream where the stream "even in the driest season has a volume of running water."[25] The site of San Carlos was also changed, both to get away from the influence of the presidio and to provide better conditions for agriculture. Again the movement was up valley. Finally, Santa Clara had not two, but three sites. The first was on the banks of the Rio Guadalupe. After three years, the mission was moved to higher ground because of floods. A final move took place after the earthquakes of 1812 and 1818.

The problem of supplying water was met at many of the missions by the construction of imposing water systems which place the Franciscans in the forefront of irrigation history in the state. At San Diego, where the problem was acute, measures were correspondingly impressive. This was the pio-

[18] *Ibid.*, Vol. III, p. 216.

[19] Z. Engelhardt, *Santa Barbara Mission* (San Francisco, 1923), p. 56.

[20] Z. Engelhardt, *Mission Nuestra Señora de la Soledad* (Santa Barbara, 1929), p. 34.

[21] Fray Francisco Durán to Governor Echeandía, January 4, 1828, Document 1049, Santa Barbara Mission Archives (SBMA). Unpublished.

[22] Duhaut-Cilly, *loc. cit.*

[23] Z. Engelhardt, *San Juan Capistrano Mission* (Los Angeles, 1922), p. 88.

[24] F. E. Green, "The San Diego Old Mission Dam and Irrigation Works," typewritten manuscript in the Junípero Serra Museum, San Diego (1933). Green reproduces the viceroyal decree which contains directions on the transfer of the site. A photostat of the original decree, dated December 17, 1773, is in the Bancroft Library, University of California.

[25] Z. Engelhardt, *San Antonio de Padua* (Santa Barbara, 1929), p. 11.

neer in water development and illustrates the order of such development at many of the missions.[26] At the first site from 1769 to 1772, water was simply baled from sump holes in the river sands. After the removal to the new site in 1773, numerous ditches were dug and sand-and-brush diversion dams were built. For over ten years, however, the new location was considered temporary because of the lack of water. After 1794, a number of small reservoirs and a masonry cistern with a mechanical lift for the garden were in operation. Finally, in 1815, the Franciscans went up the San Diego Valley and on an exposure of porphyry trap rock at the head of the gorge, six miles from the mission, they built a solid masonry dam, 224 feet long and 12 feet thick. "The site chosen for this dam," writes Green, "is ideal for a low dam for diverting the small summer flow of the stream because of its excellent foundations, least hazards from winter floods, long forebay, and approach through a wide, flat, sandy channel."[27] Apparently no extensive use was made of underground waters from the river gravels, a practice that was to become common in southern California only with irrigation settlement by the water companies in the late nineteenth century. Without pumping plants and without adequate storage facilities, the missions were still plagued by serious shortages.

Other missions whose water systems deserve particular mention are Ventura, Santa Barbara, and San Antonio. The first two are along the Santa Barbara coast which, it will be recalled, was passed over for mission sites by Font because of the lack of water. The great stone aqueduct of Mission San Buenaventura was seven miles long; it was finally demolished by the floods and landslides of 1866–71. At Santa Barbara, the Mission Historical Park contains portions of the once extensive waterworks system—a filter house, a grist mill, sections of the aqueducts, and two reservoirs. The larger reservoir built in 1806 is of masonry, 125 feet square and seven feet deep, and is still used as part of the city water system. A dam, built in 1807, is located in the Santa Barbara Botanical Garden, one and one-half miles up the canyon. The water problem along the Santa Barbara coast was not as acute as that at San Diego, and the report of Santa Barbara Mission for 1822 notes that irrigation was not necessary and the mission did not need to possess "wonderful irrigation ditches" like San Diego.[28]

One of the more complete irrigation systems was located at San Antonio Mission (Fig. 5). Here, two ditches brought water to the area around the mission. The first ditch tapped San Miguel Creek (now Mission Creek) and supplied the garden and vineyard. The second ditch led from a dam on the San Antonio River, a little more than three miles above the mission, and added water to irrigate the plain in front of the mission. The size and construction of the ditches is shown in Figure 5. From the ditches, irrigation water was carried to the fields and vineyards in clay pipes laid on a bed of tile and covered with shale and boulders. A number of wells and reservoirs are also shown. Remnants of most of these features still exist on the mission grounds.

Another aspect of water utilization was the water-operated grist mill. This was a late feature at the missions and appears in its most advanced form with the construction of a mill for Mission San Gabriel in 1816. The mill is fully described by Cleland[29] and can be taken as typical of later mills at other missions, such as Santa Barbara and Santa Inés. The mill was located at the junction of two arroyos about two miles above the mission. The water that turned the wheel passed into a pond which was drained at its lower end. The flow from the pond was used in turn to provide power for a sawmill, tannery, and wool works farther down the slope. The mill itself was a two-story building with the roof and walls of masonry and adobe brick. It was used to grind wheat and corn for seven years, after which it was abandoned in favor of a new mill built opposite the mission compound.

Mission Indians

One of the more definite requirements for a mission's location was the presence of a large Indian population. As it turned out, certain features about Indian settlement and economy were to facilitate greatly the work of the missionaries. In the first place, a

[26] Green, *op. cit.*, pp. 14–17.
[27] *Ibid.*, p. 31.
[28] Engelhardt, Santa Barbara, *op. cit.*, p. 116.
[29] R. G. Cleland, *El Molino Viejo* (San Marino, Calif., 1950).

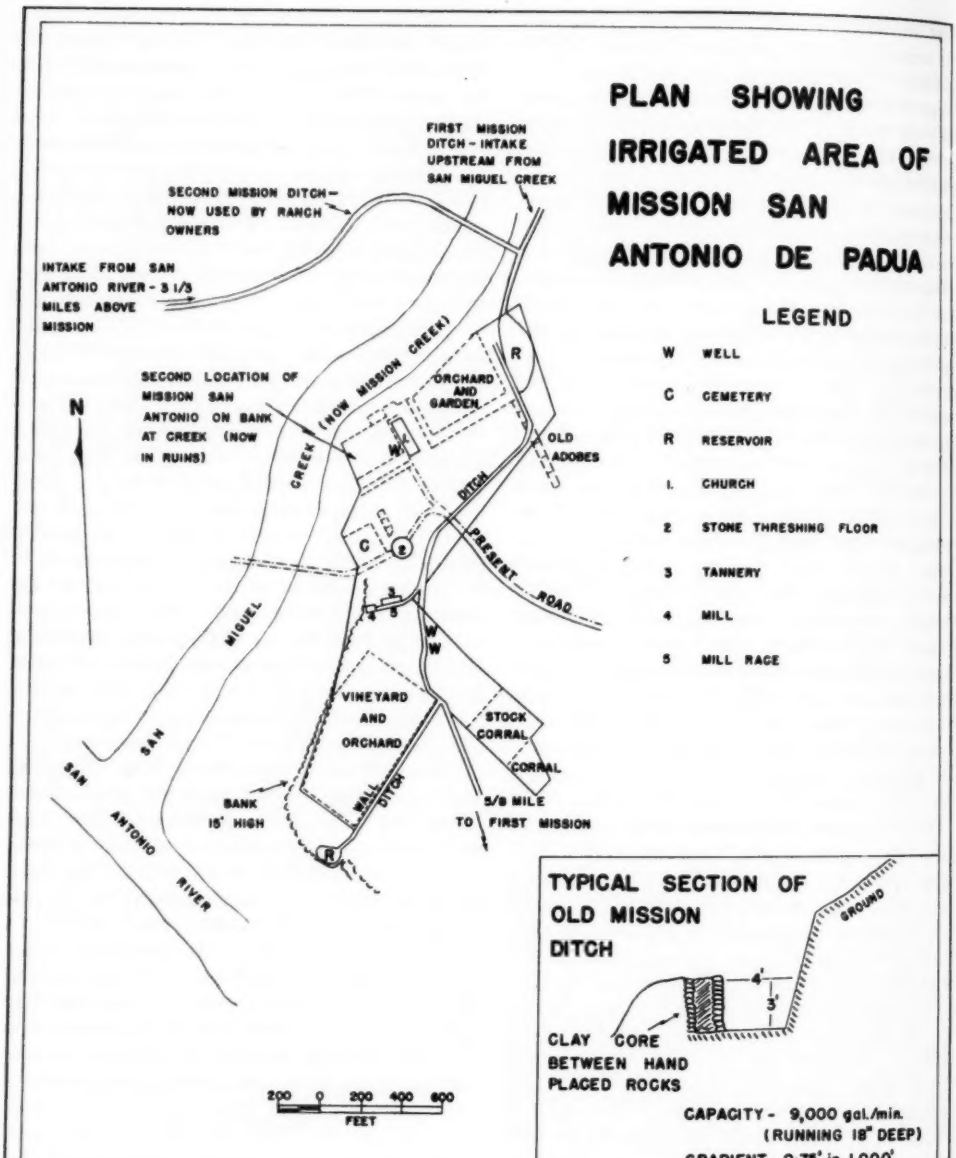


Fig. 5. Mission San Antonio irrigation system. This was one of the more complete irrigation systems among the misions and illustrates how ditches were used to tap the flow of different creeks. From the ditches, water was carried to the fields in clay pipes laid on a bed of tile and covered with shale and boulders. From F. R. Smith.

remarkably large proportion of the pre-Columbian population of what is now the United States was in California, probably as much as thirty percent. Perhaps half of this number was located in the coastal areas where missions were to be established (Fig. 2). These Indians were a gathering and fishing people, characteristically quiet and peaceful, and receptive to contact with the missionaries. In addition, their settlements were remarkably stable. They lived in rancherias or villages whose territories were well defined and rarely transgressed. Their principal food was the acorn, but other plant foods were also available in quantity. Among the Salinan Indians, for example, an incomplete listing contains six kinds of acorns, three grasses, three clovers, six berries, two pine nuts, wild oats, buckeye, sunflower, chia, sages, grapes, prickly pears, and yucca. In addition, every available variety of fish, reptile, bird, and many mammals was eaten.[30] This wealth of food materials helps to explain the density and distribution of the aboriginal population.

In some of the coastal areas, particularly along the Santa Barbara coast, rich marine resources were effectively utilized. Here the Chumash groups, with their plank canoes, fish hooks, and harpoons, were recognized as the most technologically advanced in the mission areas. Governor Fages, in his description of the Indians around Mission San Buenaventura, noted that it was the opinion of "all the early discoverers of the country and those who settled there later, that the Indians from this locality . . . are active and industrious . . . and of a quality between *gente de razón* and other Indians."[31]

However, not all of the Indian groups were well provided with food nor were they all submissive to the Franciscans. The Diegueños were particularly hostile. Although the first mission was established at San Diego, Indian resistance persisted here longer than anywhere else. The only instance of Franciscan martyrdom occurred here. The difficulties of the missionaries were bound up with the inability of the area to provide sufficient sustenance. Father Crespi noted this when the mission was being founded. "The territory," he wrote, "is more sterile than any we have seen and there is a lack of food for the poor wretched pagans. Even the daily bread . . . the maguey plant . . . is wanting."[32] It was not until after the establishment of irrigation, already noted, that some sort of regularity in food production was achieved.

Population and Production

Once established in an area well supplied with Indian villages, the mission introduced a new economy and a new way of life. The progress that was made, both material and spiritual, is well documented in the reports left behind by the missionaries. The most valuable of these is the *informe anual.* Drawn up by the mission at the close of each year, beginning in 1783, the report contained the following information: location of the mission, distance from the next mission, names of missionaries in charge, number of baptisms, deaths and marriages from the date of founding; number of neophytes; number of cattle, sheep, goats, horses, mules, and swine; number of *fanegas*[33] of wheat, barley, corn, beans, peas, lentils, and *garbanzos,* sown and harvested. The reports were returned to the Father President who made a general summary from them in triplicate. One copy was sent to the governor of California who sent it on to the viceroy in Mexico. Another copy went to the College of San Fernando, Franciscan headquarters in Mexico. The third copy was kept in California in the archives of the Father President. It is one of the more fortunate features of mission history that virtually complete files of the original reports are still in California in the mission archives at Santa Barbara. These reports have been used in compiling most of the information that follows.

A second series of reports was issued biennially beginning in 1795. These contained slightly different information for population, such as the number of male and female Indians and the number of boys and girls under nine years of age at each mission. A third kind of *informes* was made out from time to time on the conditions of the missions and the Indian neophytes. These contain particularly

[30] A. L. Kroeber, *Handbook of the Indians of California* (Washington, 1925), p. 547.

[31] Z. Engelhardt, *San Buenaventura* (Santa Barbara, 1930), p. 19.

[32] Engelhardt, San Diego, *loc. cit.*

[33] The Old Mexican measure of the *fanega* is equivalent to 100 pounds.

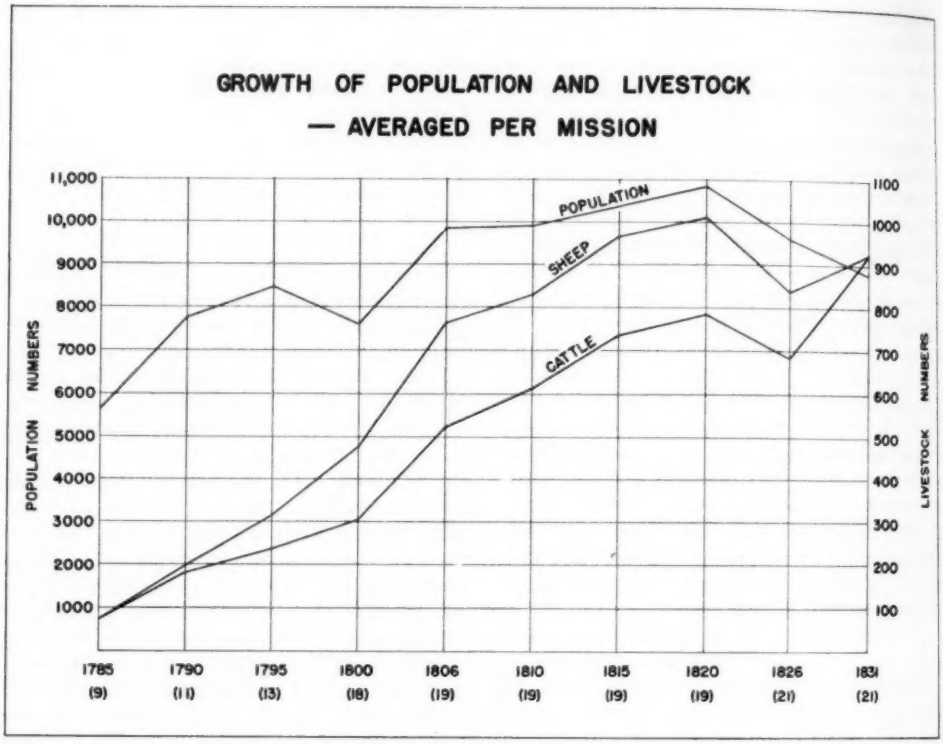


Fig. 6. Population and livestock numbers at the missions. Figures shown are the total numbers divided by the number of missions (indicated in brackets under the different years). The correspondence in the items shown is obvious; so is the peak that represents the climax in mission activity around 1820. The dip in average population from 1795 to 1800 is the result of the founding of a large number of missions in this period. The number of missions is shown in parentheses. Data from the *informes*.

useful information on land utilization, extent of mission lands, and problems of agricultural production. Unfortunately, the originals of many of these were lost both in Mexico City and in San Francisco, but copies of most of them had been made and these are available. The earlier editions of Engelhardt's "Missions and Missionaries"[34] contain many geographically useful items from these reports copied before their destruction in San Francisco in 1906.

Total baptisms at the missions from the time of their founding to secularization were more than 98,000, of which about 5,000 were probably Spanish. The largest number of neophytes lived at the missions in 1821. This is followed by a slight decline to 1831. The growth per mission (Fig. 6) shows the same general features. The peak around 1820 is evident. The dip from 1795 to 1800 is a result of the founding of a fairly large number of missions—five—in 1797 and 1798. The most rapid population growth is in the decade 1795 to 1806, by which time 19 of the 21 missions had been founded. Numbers grew from 11,025 to 18,327, an increase of 70 percent. This represents the golden age of mission expansion. The decade 1811–20, in turn, was the golden age of total growth. During these years, one-third of the missions attained their maximum populations.

When numbers are compared from mission to mission, interesting regional variations appear. In Figure 7, the missions are arranged from south to north, from San Diego to San Francisco Solano. The graph is dominated

[34] Z. Engelhardt, *The Missions and Missionaries of California* (4 vols.; San Francisco, 1908–1915).

FIG. 7. Population and livestock at the missions. Numbers are averaged according to the number of years each mission was in operation. The graphs are dominated by the four southern missions (1–4) and two in the north (17, 18). Mission names are shown in Figure 2. Data from the informes.

by the four southern missions. The first three missions in average numbers are here—San Luís Rey, San Diego, and San Gabriel. The high numbers at San Diego are surprising in view of the difficulties the mission faced. Growth was slow and numbers small from 1769 to the late 1790's. The achievements in irrigation, already noted, help to explain later growth. Both San Diego and San Luís Rey reached their population peaks after 1820.

Among the missions to the north two stand out from the rest, Santa Clara and San José. The advantages of site and location for both have already been noted. San Antonio with its irrigation system was also able to support larger numbers than the missions around it. Little can be said of the apparently impressive positions of San Rafael and San Francisco Solano since they were in operation for such a short time. (San Rafael was founded in 1817 and San Francisco Solano in 1823.) The small numbers at places like Soledad, San Carlos, and Santa Cruz are in line with deficiencies already indicated for these locations.

THE MISSION ECONOMY

A New Landscape

The founding of a mission not only heralded new population conditions; it also set in motion a series of changes each of which made its contribution to the formation of a new landscape. Most evident were the buildings themselves. "The architecture," writes Smythe, "harmonized with the landscape, for it was the architecture of Spain in a landscape resembling Spain."[35] In a general way this is true. The building styles were Spanish in origin, and basic features, such as the gently slanting roof of burnt tile, the thick adobe walls, and the extensive use of arcades, were repeated from one mission to another. But the mission buildings also displayed new features adapted to their function of caring for large Indian populations and reflecting the conditions imposed by particular sites and available building material. As Baer points out, the mission plan was functional.[36] Defense and control were important in the missionary program to make the Indian loyal and self-supporting. Hence, the mission compound took on new significance. It provided living and working space and the four walls made it a fort. The quadrangle is a central feature of almost all mission plans. One exception is La Purísima, where the buildings occupy one long row at the foot of a hill. In other places the quadrangle, although planned, was never completed.

The construction of mission buildings was greatly influenced by the availability of materials. Good building materials were not always at hand in California, and this lack, together with the lack of skilled workmen, helped bring about an individual architectural style, one that Baer calls "Californian Franciscan." This style evolved over a period of time, and Baer recognizes a number of "periods" in building, based on construction materials.[37]

In the first of these, the buildings were of wood. Posts, usually of pine, were set together, daubed with clay and coated with a wash of lime. Roofs were made of poles covered with layers of rushes and clay. This "wattle-and-daub" construction was eventually replaced by the first permanent buildings. Adobe was now used for the walls. As stone was usually not available, adobe, particularly in the form of sun-dried bricks, became the most common building material. Eventually, too, tiles replaced the rushes on the roofs. A third stage appeared with the construction of churches of stone, usually a soft sandstone that was easy to work. The more prosperous missions, such as San Gabriel, Santa Barbara, and San Buenaventura, used a combination of stone and burnt brick. The church at San Juan Capistrano was a unique structure in that stone was used throughout—for the interior, the walls, and even the roof.

The changes in the mission buildings reflected, in turn, the changes which were taking place on the land. A series of transformations accompanied agriculture. Usually, before anything else, a small vegetable garden was planted and irrigated by hand. Grain production followed, also on small plots of land. The small plots were followed by larger fields. And growing most rapidly and extending beyond the fields to the limits of the mission domain were the livestock ranches.

[35] W. E. Smythe, *History of San Diego* (2 vols.; San Diego, 1908), Vol. 1, pp. 63–64.

[36] K. Baer, *Architecture of the California Missions* (Berkeley and Los Angeles, 1958), p. 42.

[37] *Ibid.*, pp. 14, 17–26.

The materials with which the changes were carried on, the cultivated crops and domesticated animals, had never before been part of the California landscape. The missionaries introduced them, and in the process brought to their new land a variety of agricultural production that has been one of its distinctive features ever since.

Grain Production

The three principal grains produced in mission fields were wheat, barley, and corn. Annual figures for each mission are available from the *informes* from 1785 onwards. These indicate the total amounts sown and produced in each year. Area of production is not indicated, but some idea of relative yields can be obtained by comparing amounts sown and harvested. Humboldt noted an average wheat return of 16 or 17:1,[38] and although his figures are based on only two years' production, they compare well with rates based on the lengthy record from 1785 on. On this basis, rates ranged from a low of 6:1 for San Carlos to better than 20:1 for San Gabriel, San Buenaventura and San José. The less favorable conditions of production in the missions along the northern coast were specifically noted by Humboldt. With the exception of Santa Cruz, the returns here were 12:1 or less. In contrast, the returns in the south, particularly from Ventura to San Juan Capistrano, were especially high. Humboldt compared the yields with other parts of the world, and found them better than in France (6:1) or Hungary (10:1) but below those of Central Mexico (24:1).[39]

Grain production varied from one part of Alta California to another, just as precipitation and temperature conditions varied. Differences in precipitation were particularly critical. The mean annual rainfall increases from south to north, with elevation, and from the leeward to the windward sides of mountains. It is less than ten inches at San Diego and about 25 inches at Sonoma. It is less than ten inches at Soledad and nearly three times this near San Antonio, a short distance away in another valley. Mean annual precipitation, however, is only one aspect of the rainfall situation. Distribution throughout the year and variations from year to year are just as critical and receive consideration below. Nevertheless, there is a general relationship between grain production and mean annual rainfall. The highest producers among the missions were located in areas receiving 15 to 20 inches of precipitation per year—San Luís Rey, San Gabriel, San José, and Santa Clara. Places like Santa Cruz (over 30 inches) and Soledad (ten inches) suffered from too much or too little rain.

The Wheat Crop

Of the three grain crops, wheat was by far the most important. It was the major crop at each of the missions. The production of barley and corn, though widespread, was generally low. There were some exceptions. For example, barley and corn were almost as important as wheat at San Luís Rey (Fig. 8), and corn was a major crop at San Gabriel. The bulk of these three crops, it will be noted, was produced in the southern missions from San Diego to San Buenaventura. In the northern missions, even among such heavy grain producers as Santa Clara and San José, corn and barley were insignificant.

That wheat was the major crop at each of the missions is shown in Figure 8. Also shown is the dominant position of the southern missions in wheat production, from San Diego to San Buenaventura. The outstanding producer was San Gabriel. After 1790 its wheat crop was regularly the largest among the missions. Its yield of 11,000 bushels in 1821 stands as the highest attained in the agricultural history of the missions. As early as 1774, Anza noticed the "promise of fair-sized crops of wheat" at San Gabriel, adding that "the fields of this crop . . . yield very abundantly, for with respect to rankness of growth, size of head and thickness of stalk, I have not seen any so good in the provinces of Sonora."[40] Anza observed that the fogs and dampness which afflicted San Carlos were absent here, and that the soil was productive. The availability of water was a major asset, and by the end of the century most of the wheat and corn was being irrigated. Father Sánchez reported in 1796 that "all the land that could be irrigated was planted in grain."[41]

[38] Humboldt, *op. cit.*, Vol. II, p. 467.
[39] *Ibid.*, pp. 475–80.

[40] H. E. Bolton (ed.), *Anza's California Expeditions* (5 vols.; Berkeley, 1930), Vol. II, pp. 110–11.
[41] Engelhardt, San Gabriel, *op. cit.*, p. 68.

FIG. 8. Grain production at the missions. Yields are averaged according to the number of years each mission was in operation. The variations between missions reflect differences in temperature and precipitation conditions. Figures for San Rafael and S. F. Solano (20, 21) are not comparable because of the shortness of the record. The *fanega* was equivalent to 100 pounds. Data from the *informes*.

But irrigation was not the complex affair here that it was in San Diego. Extra ditches led water to areas near the mission quadrangle, but production was a good deal more far-flung than this.

Webb indicates 12 to 15 grain ranches in the agricultural organization of San Gabriel.[42] The most distant were at San Gorgonio (Beaumont) and San Bernardino. These were undoubtedly stock ranges, but others, such as La Puente and Santa Anita, produced a large share of the grain. The latter retained its position in agricultural production after secularization and was cited by Robinson as "one of the fairy spots to be met with so often in California."[43] In his report of 1827, Father Sánchez noted a part of the mission land "that is sufficiently humid to depend on the rains" and mentions the Santa Ana and San Bernardino areas as having water available in the same way as areas closer to the mission, "by reason of water ditches" which depended on rain for their supply.[44]

Almost equal in production to San Gabriel were the missions of San Luis Rey and Santa Clara. Little need be said about San Luis Rey except that it represents one of the later establishments (1798) and duplicates many of the conditions at San Gabriel. At Santa Clara, the advantages of the site, particularly in soil and water supply, have already been noted. Vancouver attributes the high production to physical conditions, all the more significant since methods were so primitive. He describes them:

> By the help of a very mean and ill-contrived plough drawn by oxen, the earth is once lightly turned over, and smoothed down by a harrow; in the month of November or December, the wheat is sown in drills, or broadcast on the even surface, and scratched in with a harrow; this is the whole of their system of husbandry, which uniformly produces them in July and August an abundant harvest.

He goes on, noting yields of twenty-five or thirty-to-one in spite of the poor method of threshing, "which is always performed in the open air by the treading of cattle."[45] Much the same observations are made by Duhaut-Cilly who describes the threshing in some detail and remarks that yields of grain could be higher with fertilization, "for until now the missionaries have sown with no other preparation than a very imperfect ploughing."[46]

In contrast to San Gabriel and Santa Clara, San Carlos represents the other extreme of grain production. Its averages for wheat are the lowest among the missions, although there is some compensation for this in the larger proportions of barley and corn that were raised. The difficulties that San Carlos had to bear are summarized in early reports. Governor Fages noted the lack of water for irrigation, a climate that was "disagreeably rough," and the location of the mission so that it was "always exposed to rough winds."[47] In the mission *informe* of 1827 the lack of summer heat is mentioned as a disadvantage. Vancouver described the wheat produced as "of an inferior quality and the return . . . by no means equal to that produced at Santa Clara." He adds poor soil quality to the list of drawbacks. "As the soil here," he wrote, "is of a light sandy nature, its productions are consequently inferior."[48] But some advantages were noted. Both Fages and La Pérouse observed that the frequent fogs were of some use in providing moisture. And it was possibly at San Carlos that potatoes were first grown. La Pérouse claims to have introduced them, declaring that this crop would "thrive very well in these light lands."[49]

Because of the importance of wheat, it has been used as the index of crop production at the missions. Changes in wheat production for eleven selected missions are shown in Figure 9. Actual production is presented for each year from 1785 to 1831 as an average figure for the group of missions. Also shown is the five-year moving average which smooths out the extreme fluctuations and serves as a basis of comparison with the first curve.

The moving average curve indicates fairly regular growth until the peak years around 1820, and then a sharp decline. The actual curve crosses this with marked peaks and

[42] E. B. Webb, *Indian Life at the Old Missions* (Los Angeles, 1952), p. 92.

[43] A. Robinson, *Life in California* (Oakland, 1947), p. 22. The well-known "Life" describes mission conditions in the late 1820's.

[44] Engelhardt, San Gabriel, *op. cit.*, p. 142.

[45] Vancouver, *op. cit.*, Vol. I, pp. 38–39.

[46] Duhaut-Cilly, *op. cit.*, p. 237.

[47] Z. Engelhardt, *Mission San Carlos Borromeo* (Santa Barbara, 1934), p. 91.

[48] Vancouver, *op. cit.*, Vol. I, p. 64.

[49] Comte J. F. G. de la Pérouse, *A Voyage Around the World* (3 vols.; London, 1797), Vol. II, p. 233.

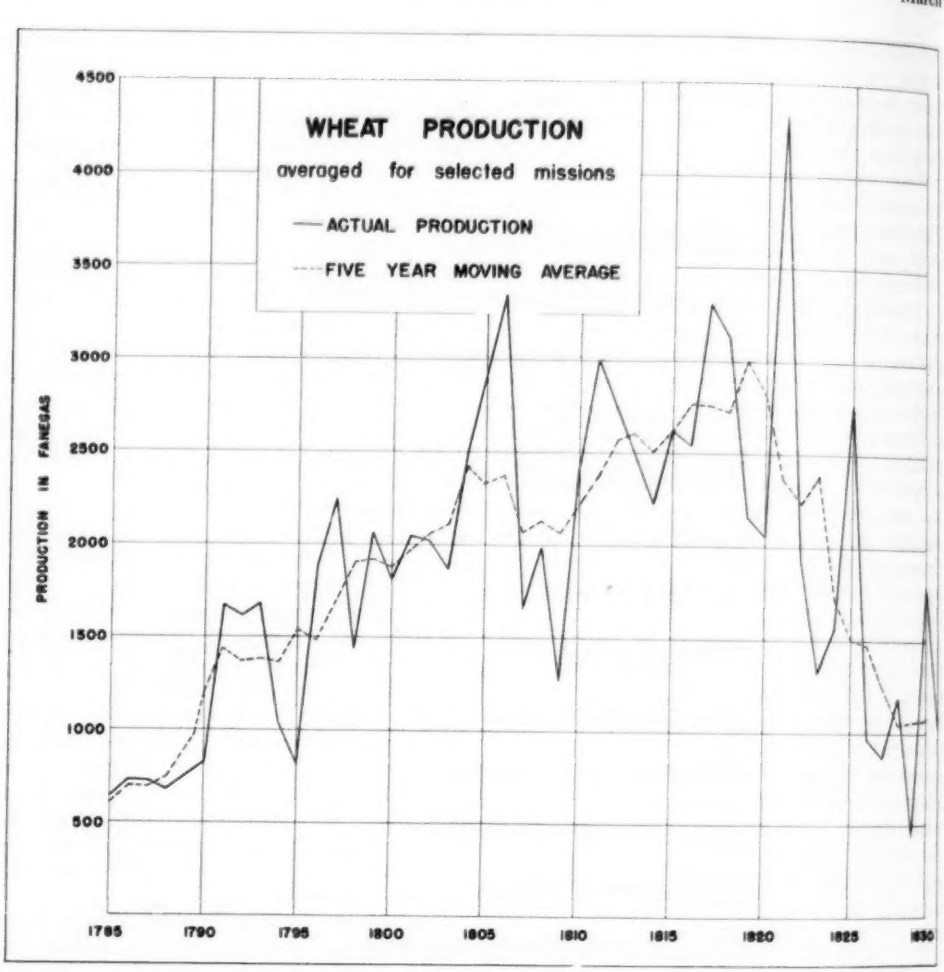


Fig. 9. Actual wheat production at the missions. For this graph, certain missions were selected and the average production of all of these was plotted for each year. The number of missions selected increases from six in 1785 to eleven in 1798 and afterwards. The curve of actual production crosses that of the moving average in marked peaks and plunges, the result of extreme variability in rainfall. Data from the *informes* and Lynch.

plunges, a vivid witness to fluctuations in production in a climate marked by extreme variability in rainfall. With few exceptions, wheat was not irrigated. Its yields were high when rain was abundant and low when it was scarce. The striking contrast between 1821 and 1823 points up the extremes.

Using crop returns at the missions as a basis, an interesting attempt has been made by Lynch to assess rainfall conditions in California.[50] His conclusions may be related to the graph of wheat production. Lynch proposes three periods of rainfall deficiency or surplus. The first, from 1782 to 1810, was one of deficiency, apparently prolonged and severe. And yet it was during this time that

[50] H. B. Lynch, "Rainfall and Stream Runoff in Southern California Since 1769," *Report to the Metropolitan Water Board*, Los Angeles, 1931.

the newly established missions needed to put their agricultural production on a firm footing. It is not surprising then that the difficulties with drought received such emphasis in the mission reports of the time. Wheat production shows three peaks in this period, including the high peak in 1806, a year in which production reached the second highest level in the agricultural history of the missions.

The second period noted by Lynch was one of surplus from 1810 to 1821. These were very good years. In Lynch's words, this was a "surplus period . . . more intense than anything in the last forty years." He compares it with the surplus in southern California from 1883 to 1893 when settlement and agricultural production combined to produce the boom of the 1880's. Wheat production was higher than at any other time. The peak years—1811, 1817, and 1821—stand out impressively. This was the golden age of agricultural production. The same period has already been described as the golden age of total population growth at the missions.

Lynch's third period, 1822–32, was one of deficiency "more severe than any since." Production dropped to the levels of the years of early growth, and just as dry years accompanied the struggles of establishment, so now they accompanied the struggle to prevent disintegration of mission lands in the face of secularization. The struggles were of no avail, however. Secularization took place, and with it, mission agriculture disappeared.

Livestock

The introduction of domestic livestock in California is usually attributed to the Rivera expedition which brought a number of animals from Baja California to San Diego in May of 1769. The herd included 123 full-grown animals, mostly cows, and 41 young ones, and was intended to provide a start for the proposed missions of San Diego and San Carlos. More animals were introduced by the Anza expeditions, 1774–76, which established land communication between Sonora and California. Later in 1776 a number of animals for breeding purposes were sent from missions in Baja California. This seems to mark the end of livestock introductions.[51]

By the end of 1776 livestock numbers were enough to provide the seven missions established at the time with their most lasting bases for settlement. In addition, stock from these missions was used to establish new herds at subsequent missions. "From this beginning," Burcham writes, "was developed the herds and flocks which formed the foundation of the domestic economy of California, both Spanish and American, for more than two-thirds of a century."[52]

The natural conditions for pastoralism in the California colony were noted by chroniclers who accompanied expeditions in the early days of mission establishment. The master chronicler of the second Anza expedition, Father Pedro Font, provides one example. It was Father Font, described by Bolton as possessing a "sharp eye for landmarks and a canny knack of telling what he saw," who first assessed the grazing lands of the San Bernardino and San Gabriel valleys, later to develop into the extensive feeding grounds of Mission San Gabriel. He described extensive areas "covered with good pasturage, both dry and green—a country well suited for sheep and goats because it is very clean, without anything which might injure the wool." At San Antonio he noted "in the range . . . an abundance of oaks, live oaks and pines and consequently plenty of pine nuts and acorns"—conditions inviting the raising of large numbers of hogs.[53]

The grazing values of the California range were such that "it constituted a forage resource of first quality and impressive magnitude."[54] The chief forage areas were the regions of oak woodland and prairie, dominated by perennial bunch grasses, chiefly needle grass, and containing large numbers of native legumes. In addition, chaparral covered extensive areas and, although a minor range resource, contained large sections in which the brush was not too dense. Here, grasses were available and provided good forage.

The association of the missions with the major forage areas was very close. Mission lands and lands used for forage were almost synony-

[51] L. T. Burcham, *California Range Land* (Sacramento, 1957), pp. 130–33, and Webb, *op. cit.*, pp. 168–69.

[52] Burcham, *op. cit.*, p. 51.

[53] Font, *op. cit.*, pp. 171, 278.

[54] Burcham, *op. cit.*, p. 92. The condition of the pristine ranges of California is described in detail by Burcham, who has also drawn up a table indicating estimated grazing capacity of the range lands for Spanish livestock.

mous terms. The actual extent of these lands is difficult to determine. The limits changed from time to time, particularly under pressure from private ranches and nearby pueblos. In some cases the limits were not actually known, particularly where large mountain stretches were involved. However, an examination of reports from the missions in the 1820's detailing land limits provides sufficient information for an estimate. Using maximum figures yields a total acreage of 11.8 million. This represents 12 percent of the total area of the state and 22 percent of the grazing capacity estimated by Burcham for California range lands. It is doubtful that mission lands were more extensive than this.[55] They may have been a good deal less.

The most important range animals were cattle and sheep. Their growth in numbers, as recorded in the *informes*, is shown in Figure 6. The slow growth before 1785 is easy to understand. Dependence upon the animals for food and lack of new animals from the outside combined to keep numbers low. But after 1785 growth was rapid, the rate in the decade 1795 to 1806 being particularly remarkable. The numbers reflect the advantages available to the livestock industry: the almost unlimited free range land, the productivity of the forage, the lack of competition from wild animals, the freedom from diseases and pests, and the low labor requirements. The rainfall conditions which plagued crop production were not as serious for the raising of livestock. The relationship between livestock numbers and the rainfall cycle already cited is not as clear as for wheat. The period of deficiency from 1782 to 1800 was also the period of most rapid growth in livestock numbers, although a small leveling off is indicated from 1806 to 1810. Growth continued in the next decade, and it may well be that the surplus of precipitation postulated for the period was a major element in the large numbers attained in the early 1820's. After this a short decline occurred, but in contrast to the population curve with which there is a good correspondence up to 1820, numbers rose again and were very high. Actually the largest numbers of both cattle and sheep occurred in 1831, each totaling approximately 200,000. Even these bear little relationship to the figures quoted by some later observers at the missions. Alfred Robinson and Duflot de Mofras, [56] for example, grossly exaggerated numbers.

Another element in the matter of numbers was the quality of the animals. No doubt their hardiness was a major factor in their multiplication, and it is obvious that little attention was paid to maintaining or improving quality. Webb cites early descriptions indicating what the cattle probably looked like.[57] They were "leggy" and thin-bodied with long curving horns and sloping hind quarters, descendants of the same Spanish stock as Texas longhorns and cattle in the Southwest. Centuries of inbreeding produced animals more or less ferocious looking, half-wild in appearance and of many colors, generally solid reds, browns, blues, or blacks. Sheep were likewise uniform over the areas of Spanish colonization. In the Spanish and Mexican periods the typical sheep of California was the *churro*, a small, scrubby animal whose meat was poor and whose wool was skimpy and coarse.[58]

Distribution of Livestock

The distribution of cattle and sheep among the missions is shown in Figure 7. Again, dominant production is in the south. Cattle numbers are particularly high from San Diego to San Buenaventura. They drop off noticeably after this and remain uniformly low. The southern dominance is not as marked for sheep, but the numbers from San Diego to San Gabriel are higher than for any other group of missions. Three of these four missions rank among the first four in sheep numbers.

Livestock numbers at each of the missions were determined by two factors, the neophyte population and the extent and condition of the mission lands. The relationship to population is apparent from Figures 6 and 7. The number of neophytes was a basic consideration in the extension of herds. Here was a main source

[55] Burcham quotes Robinson's estimate that the mission lands made up 17 percent of the total area of the state.

[56] E. Duflot de Mofras, *Travels on the Pacific Coast*, ed. by M. E. Wilbur (2 vols.; Santa Ana, Calif., 1937). This material was first published in Duflot de Mofras, *Exploration du Territoire de L'Oregon, des Californies et de la mer Vermeille . . . pendant les années 1840–1842* (Paris, 1844).

[57] Webb, *op. cit.*, pp. 174–75.

[58] Burcham, *op. cit.*, p. 59.

of food and of industrial materials. From livestock came meat; hides for clothing, shoes, rope, and harness; fats for soap and lubricants; and wool for weaving yarn. Increases in the herds meant more demands on labor, but this was never a critical item. The presence at each mission of herds of cattle, sheep, and horses and smaller numbers of mules, goats, and pigs would call for employment but only on a modest scale.

The availability of grazing land was the most important factor governing the size of the herds. The missions with the large land holdings were the main livestock raisers. Calculations on land limits of the missions indicate holdings of over one thousand square miles at each of the following: San Diego, San Luís Rey, San Juan Capistrano, San Gabriel, San Miguel, San Antonio, and San José. With the exception of San Antonio and to a lesser extent San Miguel, both with large areas in rough mountainous country, these were the main livestock raisers. In some cases, such as San Buenaventura where the lands covered an area of about six hundred square miles, the quality of some of the pasture helped to compensate for the smaller areas available. In contrast, the areas available to San Diego were very extensive, probably over two thousand square miles and second in size only to those at San Gabriel. This vast area was needed because of the low carrying capacity of the ranges.

Each of the missions, regardless of the size of its lands, valued them highly. Among the most frequent references in the many mission reports are those which emphasize the lack of pasture land and the necessity of resisting encroachments from private ranchers or even from other missions. The attitude of the mission fathers is summed up in a letter by Father Fernando Martín at San Diego, protesting the demand that one of the mission ranches, which was used as supplementary grazing ground, be transferred to private ownership. He wrote:

> To the Indians belong the lands and the site where they and their fathers . . . had been born, for nature has given it to them . . . . Who doubts that it is the Mission which suffers if that site is given away? In this year of drought [1829], when there is no pasturage for the sheep, where shall they be placed? Which localities has the Mission to maintain seventeen thousand sheep during the five months of cold weather? Certain places are set aside on which to keep the livestock of the Mission, in such seasons of drought and . . . [this] . . . is one of them.[59]

Even at San Luís Rey, where pasture was superior to that at San Diego, aridity was scored as a reason for not giving up land. In a report to the Imperial Commissioner from Mexico, it was pointed out that the land was "to a great extent sterile . . . for that reason the livestock needs it all; indeed in years when there is little rain, the stock dies for want of pasture."[60] Despite these difficulties San Luís led the way in livestock increases in the 1820's. Its cattle population of 27,500 in 1832 was the largest ever recorded for any mission.

Mission Ranches

Each of the missions had a number of ranches for the raising of livestock, and in the post-mission period it is the ranch which remains as the one persistent element of the California scene. Cleland describes it as "one of the few enduring legacies from Mexico and Spain."[61] A particular ranch might be located near the mission compound or a long distance away. It might be a ranch devoted to the raising of a single type or a combination of livestock. Sometimes it was used for crop production. In a few instances it grew important enough to become an *asistencia* to the main mission. Then it became a miniature mission with its own chapel and compound and a sizable resident population.

The number of ranches belonging to a mission usually reflected the importance of its livestock industry. Mission San Luís Rey at one time had more than twenty ranches scattered over an area of approximately 1,200 square miles. The most important of these were Rancho San Juan, eight miles east of the mission, used exclusively for cattle; another rancho, over forty miles eastward, used only for sheep; Rancho Las Pulgas, ten miles to the north, and Rancho San Jacinto, thirty miles to the northeast, both used only for cattle; and the mixed ranches of Santa Margarita, three miles to the northwest, San Pedro (later known as Las Flores), five miles beyond this,

---

[59] Engelhardt, San Diego, *op. cit.*, pp. 226–27.

[60] Z. Engelhardt, *San Luis Rey Mission* (San Francisco, 1921), p. 41.

[61] R. G. Cleland, *The Cattle on a Thousand Hills* (San Marino, Calif., 1941), p. 33.

and San Antonio de Pala, eighteen miles to the northeast. The latter provided one of the best examples in the colony of a mission station or *asistencia*. In a report to the governor, dated December 22, 1827, it is described as containing its own chapel, dwellings, and granaries; a "few fields" of wheat, corn, beans, and garbanzos, a vineyard, and fruit and olive trees which were irrigated.[62] Livestock of all kinds was apparently kept. The Las Flores rancho was described by Robinson in 1829 as one of the cattle establishments of Mission San Luís Rey. "It is situated," he wrote, "on an eminence commanding a view of the ocean . . . and overlooking an adjacent level, extending for miles around, covered with thousands of animals grazing."[63] A few years after secularization, Las Flores, together with Santa Margarita, was to comprise the largest patented land grant issued in California, a total holding of 133,441 acres.

The ranchlands at San Gabriel were the most extensive in the whole mission chain, extending over an area of approximately 2,100 square miles. The number of ranches was correspondingly high. Duflot de Mofras indicates a total of seventeen ranches "for large animals" and fifteen "for smaller animals."[64] The extensive ranching domain of Mission San Gabriel made it particularly susceptible to demands that parts of its land be released for grants to individuals. In an application for such a grant in 1791, one Lt. José Francisco de Ortega pointed out the vast extent of the holdings. "Toward the east," he wrote, "where the cattle and horses of this Mission have their pastures, there is a vast stretch of water, soil, pastures, timber, etc. To the south, likewise are the thousands of cattle and the horses of the Mission; on fertile lands . . . to the north, along the foot of the Sierra, are the ranches for the sheep."[65] A fairly large number of these petitions for land were granted. Bancroft contains a list of some of them, each one delivered over strong protests by the mission.[66] The largest was the Los Nietos grant near La Puente which was in contention from 1784 into the next century. It carried over one thousand cattle and at its disposition by the United States Land Commission in the mid-nineteenth century, it was divided into five tracts with a total area of 140,000 acres. In a letter dated March, 1796, the further granting of land in the La Puente area under pressure from a nearby rancher is protested by Father Sánchez. "At La Puente," he wrote, "the mission had three thousand head of cattle and more than four thousand sheep, besides the horses. They needed the pasture and the water . . . the Indians are poor indeed while Nieto [the rancher] is not poor; he has more horses than the mission . . . [and] . . . numerous cattle."[67]

San Gabriel's difficulties were heightened by the presence nearby of the pueblo of Los Angeles. The use of mission land for the grazing of pueblo livestock was a regular occurrence and brought a steady flow of complaints from the mission fathers. With secularization, the pueblo and its inhabitants no longer had the mission to contend with, and livestock operations were expanded accordingly. In the words of Duflot de Mofras, "the Pueblo of Los Angeles is extremely rich, for the spoils of the neighbouring missions have fallen into the hands of the local inhabitants; and within an area of fifteen or twenty square leagues, local residents own over 80,000 cattle, 25,000 horses and 10,000 sheep."[68] The figures are exaggerated but the new prosperity was real.

In the north, the largest livestock numbers regularly occurred at Mission San José. The extent of the mission lands, its ranches, and the Indian settlements are shown on a "topographic plan" drawn in 1824.[69] The total area over which livestock ranged was approximately 2,000 square miles, making the range lands comparable in size to those of San Gabriel. In a description of the lands written by Father Francisco Durán in January of 1828, three chief ranching areas are described.[70] These can be located on the "plan." The first is a tongue of land northwest of the mission, with Rancho San Lorenzo in the center. Here

[62] Engelhardt, San Luís Rey, *op. cit.*, pp. 51–52.
[63] Robinson, *op. cit.*, p. 18.
[64] Duflot de Mofras, *op. cit.*, Vol. I, p. 182.
[65] Engelhardt, San Gabriel, *op. cit.*, p. 64.
[66] H. H. Bancroft, *History of California* (7 vols.; San Francisco, 1884–1890), Vol. I, pp. 661–63.
[67] Engelhardt, San Gabriel, *op. cit.*, pp. 66–67.
[68] Duflot de Mofras, *op. cit.*, Vol. I, p. 186.
[69] The "plan" is reproduced in California Division of Mines, *Geologic Guidebook of the San Francisco Bay Counties* (San Francisco, 1951), p. 36. The original is in the Bancroft Library, University of California.
[70] Durán to Echeandía, *op. cit.*, fn. 21.

were kept "nine thousand head of cattle, seven thousand sheep, and handfuls of tame horses." The arroyo at San Lorenzo provided the chief source of water. The second area is one labelled "Valle de San José," the present Livermore Valley. In 1818, it maintained "nine thousand head of cattle and ten thousand lambs." The regular sheep grazing grounds were in a third area, the "Roblar de los Chupcanes." The report concludes: "Outside of these three places, all the territory adjacent to this mission, fifteen leagues to the north, northeast and east . . . lacks water . . . and is therefore useless for sites or ranches." The limitation extended to the mountainous areas shown on the 1824 sketch.

Mission Trade

The raising of cattle was not only the most important agricultural activity on the mission lands; it was also the basis of industry and trade. The chief value of cattle was for hides and tallow. Remains of tanning vats, built of brick or stone with mortar, are still to be seen at San Gabriel and San Antonio. Other missions, such as San Juan Capistrano and La Purísima, still retain the ruins of tallow and soap vats, once part of the whole complex of workshops with dyeing, weaving, and candle-making shops nearby. Usually the part of the tallow nearest the hide was scraped off and used in household and workshop. The interior fat, probably twenty-five to one hundred pounds per animal, was a principal staple in the export trade. Hides for export were treated in brine and dried on the beach to await the ships. In addition to tanning, hides were also cured through salting and drying, and in this form served a variety of uses from covers for cot and bed frames to the ubiquitous rawhide strips.

The trade in hides and tallow seems to have begun around 1800. Unfortunately, figures for the movement are lacking. Under the Spaniards the shipping trade was limited to Spanish vessels, or after 1816, subject to excessive tariffs and duties. Any trade carried on with non-Spanish vessels would hardly be publicized. In Bancroft's words: ". . . the missionaries and settlers bartered . . . products for miscellaneous articles and money brought by the Americans whenever the opportunity presented itself, sometimes with and sometimes without the consent or knowledge of the local authorities."[71]

The Spanish trade, particularly with Lima, was considerable. Bancroft notes the arrival of two Lima ships in 1817, loaded with goods to be exchanged for tallow, of which "there was a surplus of 1,250 tons in the country."[72] Official policy was that the mission should provide fifty tons of tallow to each presidio to be used in trade. The missions, however, preferred to trade directly with Lima ships or to ship tallow to Mexico. Little tallow reached the American market. Boston merchants who picked up tallow usually traded this to the local representative of a Mexican firm in exchange for hides.[73]

Hides were the main interest of the Americans, and beginning in the 1820's, hide ships went from port to port along the coast, collecting hides and storing them in sheds preparatory to shipment. There were only three all-year ports on the coast—San Francisco, Monterey, and San Diego—but other places had open roadsteads which could be used in summer. At San Diego, trading reached an early heyday based not on cattle hides but on otter skins. Smythe describes activity in the spring of 1803 as New England ships tried to smuggle skins at a time when "virtually the whole population had skins to sell, openly or covertly, and the Commandant had a collection of about a thousand confiscated items."[74] Later on, both the port at San Diego and the roadstead at San Juan Capistrano served as major storage areas for cattle hides to be picked up by American boats. Production from the vast lands of San Luís Rey regularly moved to San Diego, although occasionally hides were allowed to be put on ships at San Juan Capistrano.

At San Gabriel, Robinson noted that the "corridor in the square is usually heaped up with hides and tallow."[75] Harrison Rogers, clerk of the Jedediah Smith expedition, made some observations on the mission's trade in December, 1826. He mentions annual mission shipments to Europe of "from twenty to twenty-five thousand dollars worth of hides

---

[71] Bancroft, *op. cit.*, Vol. II, p. 183.
[72] *Ibid.*, p. 282.
[73] R. A. Rydell, *Cape Horn to the Pacific* (Berkeley and Los Angeles, 1952), p. 55.
[74] Smythe, *op. cit.*, Vol. I, p. 107.
[75] Robinson, *op. cit.*, p. 22.

and tallow and about twenty thousand dollars worth of soap" and describes negotiations between an American brig owner and one of the mission priests for hides, tallow, and soap.[76]

The peak of the California hide trade was from 1826 to 1848 when it is estimated that California shipped about one million hides, most of which were carried in New England vessels and sold to tanneries in Boston and New York. In the same period, more than sixty million tons of tallow were shipped.[77] In the 1820's the missions were the principal customers of American ships, exchanging hides and tallow for sugar, tea, coffee, rum, furniture, clothing, and other products. Duhaut-Cilly noted a total mission exportation in 1827 of 1,200 tons of goods (four ships of three hundred tons), mostly hides and tallow, plus a few other skins and some grain. He lists 40,000 cattle hides and 45,000 *arrobas* of tallow.[78]

The increased trade was the direct outgrowth of the revolt in Mexico in 1820 and the consequent stoppage in the flow of goods to California. The missions were left completely to themselves and forced to turn to more trade to obtain supplies they needed. After 1816, the Spaniards, eager to purchase any cargo, Spanish or foreign, for the colony, opened trade to foreign ships. The growth of the trade made land even more valuable than it had been before. The pressure to increase private holdings mounted, culminating in the secularization of the missions and of mission property in the early 1830's. After this, the missions disappeared as traders.

Horticulture

Orchards, vineyards, and gardens provided the remaining element of land utilization at the missions. The planting of fruit and vegetables was invariably the first stage in agricultural development, and the watering by hand of the vegetable gardens marked the rude beginnings of what were to become in some instances impressive irrigation systems. Unfortunately, with the exception of some vegetable crops — beans, *garbanzos*, *chicharos* (chick peas), and *habas* (horse beans) — garden production was not reported in the *informes*. There are no statistics for olive and vineyard acreages or yields. But from correspondence, trade accounts, and inventories made at the time of secularization, some idea can be obtained of the extent and significance of this type of land use.

In the first place, horticulture and gardening were confined almost entirely to the missions. "Hardly a colonist from San Diego to Sonora planted a fruit tree," writes Hittell.[79] The gardens themselves were located close to the mission buildings and were usually fenced with adobe walls or cactus hedges, with wide ditches around them to keep out cattle. In addition, the Indian neophytes in many places maintained their own gardens, devoting them primarily to vegetables. At San Buenaventura, for example, Robinson noted "many small gardens . . . along the margin of the river San Buenaventura . . . belonging to the Indians where they have fruits and vegetables, which are taken to the town and disposed of."[80]

San Buenaventura also provides a good example of the possibilities of production in the mission gardens. Vancouver marvelled at the quality, quantity, and variety of plants that he saw here. He listed the following: apples, pears, plums, figs, oranges, grapes, peaches, pomegranates, bananas, cocoanuts, sugar cane, indigo, kitchen herbs, and roots.[81] Granted that some of this production was exotic rather than staple, it indicates the impressive array that was present in some mission gardens. At San Gabriel, Rogers noted that the mission possessed "large vineyards, apple and peach orchards and some fig trees."[82] Hugo Reid provides a graphic description of the way in which mission gardens were laid out. Referring to the work of Father Zalvidea who came to Mission San Gabriel in 1806, he writes:

> It was he who planted the large vineyards, intersected with fine walks, shaded by fruit trees of every description, and rendered still more lovely

[76] H. C. Dale, *The Ashley-Smith Explorations and the Discovery of a Central Route to the Pacific, 1822–1829* (Cleveland, 1918), pp. 210–11.

[77] Rydell, *loc. cit.*

[78] Duhaut-Cilly, *op. cit.*, p. 320. The figures must be accepted with a good deal of reservation. The *arroba* was approximately 25 pounds.

[79] T. H. Hittell, *History of California* (4 vols.; San Francisco, 1885–1897), Vol. II, p. 474.

[80] Robinson, *op. cit.*, p. 32.

[81] Vancouver, *op. cit.*, Vol. II, p. 229.

[82] Dale, *op. cit.*, p. 200.

by shrubs interspersed between . . . who laid out the orange garden, fruit and olive orchards . . . built the mill and dam . . . made fences of tunas (*Cactus opuntia*) round the fields . . . made hedges of rose bushes . . . planted trees in the mission square, with a flower garden and sun dial in the centre . . . .[83]

The typical layout of mission grounds is shown in the Santa Barbara plan (Fig. 10). Here the main horticultural areas are outlined and their relationship to irrigation facilities and to the workshops around the mission is shown. The small orchard east of the mission was approximately three acres in size and was devoted to olives, figs, apricots, and some citrus. The large plot close to the Indian village comprised about sixteen acres and was used for vegetables and the main olive orchard. In addition, a plot of nine acres on the bank of nearby San José Creek contained the largest vineyards.

Vineyards and olive groves were the most important features of mission horticulture. Together with the mission buildings, they provided the most obvious display of the new culture introduced by the Franciscans. The first vineyard cuttings were introduced from Baja California to San Diego and San Juan Capistrano in the spring of 1779.[84] These were first fruited in 1781 or 1782, and wine-making was probably being carried on in 1783 or 1784. By the 1820's, grapes were being cultivated as far north as San Francisco Solano. Vineyards ranged in size from the 170 acres attributed to San Gabriel to less than one at many of the missions. San Gabriel must occupy the premier position in horticulture not only because of the size of its operations, but also because its famous "Vina Madre" was used as a source of vineyard plantings throughout southern California. Duhaut-Cilly noted that the mission's wealth "consisted of immense herds and fine vines producing very good wine."[85] Robinson declares that San Gabriel annually produced "six hundred barrels of wine and two hundred of brandy" from its vineyards.[86]

Vineyard production is also mentioned prominently at Santa Barbara, La Purísima,

FIG. 10. *Ground plan of Mission Santa Barbara, illustrating the common layout of mission grounds. The main agricultural areas are indicated, and their relationship to irrigation facilities and to the workshops around the mission is shown. The total acreage allowed this mission by the U. S. Land Commission Office in 1861 is also indicated.*

San Luís Obispo, San Miguel, and San José, each of which had more than one site for growing grapes. The vineyard at San Luís Obispo was second in size (48 acres) to San Gabriel. La Purísima had two vineyards, one of 17 acres and one of five. In a letter from the mission in 1810, Father Payeras describes the division of the smaller vineyard with a nearby Spanish family in return for the work of caring for the vines. He wrote of two vineyards from which "we have old and new [wine] in abundance, made here after the Majorcan method in new wine presses."[87] The wine presses from which juice was trodden from the grapes still survive at some of the missions. Webb lists the following: two at San Antonio, one at San Miguel, one on the La Purísima lands, a wine cellar at San Fernando, an adobe building at San Gabriel which once contained three wine presses, and four masonry vat-like structures at San Juan Capistrano which may have been wine presses.[88]

The growing of olives was not nearly as extensive as that of grapes. Groves are specifically mentioned at only a few missions, but this is not the best guide to their extent. Engelhardt declares that by 1802 vineyards

[83] Hugo Reid, *The Indians of Los Angeles County* (Los Angeles, 1926), pp. 53–54. Originally published in 1852.
[84] Webb, *op. cit.*, p. 95.
[85] Duhaut-Cilly, *op. cit.*, p. 248.
[86] Robinson, *op. cit.*, p. 22.
[87] Z. Engelhardt, *Mission La Purísima Concepción* (Santa Barbara, 1932), p. 26.
[88] Webb, *op. cit.*, p. 220.

were located at half of the missions, but olive groves only at some.[89] Webb believes that no olives were planted before 1785–90, since olive mills and screw presses were not set up before 1790–95. Bancroft states that olive groves were in production at "many missions" by 1810 and by 1818 they were an "assured success" in agricultural production.[90]

The first plantings of olives, probably from seed, were made at San Diego. In 1794, a grove, with its surrounding adobe wall, is prominently mentioned and described. In the Biennial Report of 1803, it is stated that ". . . in some missions they have begun to harvest olives and at San Diego they have already made some very good olive oil."[91] This is the first specific reference to olive oil in the mission accounts. The groves at San Diego became the source of subsequent plantings at the other missions, and some of the ancient trees still here may well deserve their fame as members of the mother orchard of all the mission olive trees in California.

Olive groves also came in for special mention at San Luís Rey and Ventura. Duhaut-Cilly refers to one of the gardens at San Luís Rey as producing "the best olives and the best wine in all California."[92] The cultivation of vineyards and olive groves at San Buenaventura is mentioned in an 1828 report as a chief feature of production at the mission. In an earlier report (1822) this was emphasized even more and the gardens were described as more productive than the grain fields.[93] Olive mills were still standing at San Diego, San Juan Capistrano, and San Fernando as late as 1885.[94]

[89] Engelhardt, Missions and Missionaries, *op. cit.*, Vol. II, p. 599.
[90] Bancroft, *op. cit.*, Vol. II, p. 418.
[91] Engelhardt, San Diego, *op. cit.*, p. 154.
[92] Duhaut-Cilly, *op. cit.*, p. 228.
[93] Engelhardt, San Luís Rey, *op. cit.*, pp. 59, 66.
[94] Webb, *op. cit.*, pp. 227–28.

CONCLUSIONS

Vineyards and olive groves, then, take their place with cattle and sheep ranches, grain fields, and vegetable gardens in highlighting the material achievements of the missions. But it should be emphasized that economic aims were not the paramount ones for the Franciscans. "The Friars came to California," writes Engelhardt, "as messengers of Christ . . . not as scientists, geographers, ethnographers, or schoolmasters, nor as philanthropists."[95] But they were all these and more. Conditions of safety and material comfort had to precede or accompany their mission work. The complaint of one of the padres at Mission San Juan Bautista puts the situation in graphic terms. In a letter to the Collector of Revenues at Monterey, he wrote, "I scarcely know what to do in these troublous times. I made the vows of a Friar Minor; instead I must manage temporalities, sow grain, raise sheep, horses and cows. I must preach, baptize, bury the dead, visit the sick, direct the carts, haul stones, lime . . . ."[96]

In many ways, in the size and in the significance of their undertaking, the achievements of the Franciscans can only be called colossal. One cannot help but marvel at the way in which a handful of men took on tasks that are now the work of many groups of specialists. Agriculture, pastoralism, horticulture, and gardening were introduced and developed by them. So was a certain amount of trade. They brought to Alta California a new set of economic activities. They created a new California. And though the mission structure disappeared with secularization, the land it had transformed continued to bear its marks. Some remain even today, reminders of the part played by the missions in the changing geography of the state.

[95] Engelhardt, Missions and Missionaries, *op. cit.*, Vol. II, p. 242.
[96] Z. Engelhardt, *Mission San Juan Bautista* (Santa Barbara, 1931), p. 38.

THE DYNAMICS OF STORE TRADING AREAS AND MARKET EQUILIBRIUM[1]

WILLIAM APPLEBAUM AND SAUL B. COHEN
Harvard University and Boston University

BUSINESSMEN and marketing geographers are interested in the trading areas of retail stores. Businessmen are interested in store trading areas because these areas contain the source of business—customers, current and potential. Knowledge of a store's trading area is essential for judicious investment in facilities, merchandise, and promotional activities to attract and serve customers.

Marketing geographers are interested in store trading areas because they are complex areal phenomena. In studying these areas, marketing geographers seek to provide scientific knowledge and specific assistance to businessmen in evaluating market area potentials. A market area is an area of demand—it includes existing or potential buyers of goods and services within clearly defined geographical limits. These areas may be trading centers and their hinterlands or metropolitan districts; or, they may correspond to political areas, such as countries, states, counties, and cities; or, they may represent the administrative units of business firms.

This paper deals with:

1. Store trading-area features and the conditions under which changes in these features occur.
2. The networks formed by store trading areas of different sizes and ranks in a given market area, and the conditions under which regroupings occur.
3. Stores of market-area saturation in relation to the dynamics of trading-area features and trading-area networks.

These topics are not individual and separate, but are interconnected. A market usually contains stores of different types and sizes. Each store has its trading area. When the market area is served by stores so located, physically planned, and merchandized that customer needs are met adequately, and use of store facilities is efficient, then the market is in balance, or saturated.

Trading areas of stores do not remain fixed in time because of a multitude of dynamic changes. Likewise, store facilities become outmoded. As new store facilities, merchandising methods, or services are introduced, new trading-area patterns emerge, resulting, in time, in new trading-area networks within a market area. This process leads to a new state of balance between retail facilities and their use.

The concepts presented here are derived and broadly generalized from retailing experience in the United States and Canada, particularly in the post-World War II era. The presentation is a description of what is currently taking place. These concepts should have special relevance to store development in other parts of the world as they too adopt mass retailing methods, achieve widespread use of the automobile, and create planned shopping centers.

Marketing geography is a highly empirical discipline, with those who work in the field acting as practical engineers. A greater number and variety of samplings will have to be made of retail store trading-area patterns and networks, in the United States and Canada and elsewhere, before clear-cut theories can be offered for universal application.

[1] Appreciation is expressed to the following for their critical reading of this manuscript: Edward A. Ackerman, Carnegie Foundation; Wroe Alderson, Alderson Associates, Inc., Philadelphia, Pennsylvania; Hans Carol, University of Cincinnati; Reavis Cox, University of Pennsylvania; Bart J. Epstein, Stop & Shop, Inc., Boston, Massachusetts; Howard L. Green, Montgomery Ward, Chicago, Illinois; Richard Hartshorne, University of Wisconsin; Edgar M. Hoover, Pittsburgh Regional Planning Association, Pittsburgh, Pennsylvania; Curt Kornblau, Super Market Institute, Chicago, Illinois; Harris Lawless, Star Market Co., Boston, Massachusetts; Theodore Levitt, Harvard University, Graduate School of Business Administration; Lawrence C. Lockley, University of Santa Clara; Malcolm P. McNair, Harvard University, Graduate School of Business Administration; Raymond E. Murphy, Clark University; Richard L. Nelson, Real Estate Research Corporation, Chicago, Illinois; Jack L. Ransome, The Kroger Co., Cincinnati, Ohio; Joseph A. Russell, University of Illinois; Grady O. Tucker, Jr., Montgomery Ward, Chicago, Illinois; and Edward L. Ullman, Meramac Basin Research Project, St. Louis, Missouri; and to Eileen Schell for her cartographic suggestions and final drawings.

TRADING AREAS

A store meets the needs and wants of customers who move about within a spatially ordered framework. This movement may be related to the place of residence, work, recreation, or other activities, but the movement occurs within a specific area. The economic function of the store is to sell goods, services, or both. The store's trading area is determined by what the store sells, the manner in which it sells, the shopping habits of customers, existing competition, and type of location.

Store trading areas are phenomena that are fixed in space for recognizable periods of time. Their boundaries are zones, not lines; the boundaries are flexible, not rigid. These boundaries change with changes in shopping-habit patterns and with changes in merchandising practices.

Store trading areas are the joint product of many simultaneously interacting factors, so numerous that they almost defy generalization. To overcome this dilemma, those who attempt to define store trading areas must either offer a specialized definition that applies to a particular store trading area, or a broad generalization that may not be very applicable to any specific store trading area.

Some of the specialized definitions use as major criteria such factors as drawing power, per capita sales, share of market, movement, population, and competition.[2] *Drawing power* refers to the percent of sales that a store obtains from a specific area in relation to its total sales. *Per capita sales* refers to sales per unit of population. *Share of market* refers to percent of total available sales that a particular store obtains. *Movement* refers to transportation in terms of travel time or distance. *Population* refers to the number of people required to support a specific store. *Competition* refers to the location of competitive facilities and the power exerted by each in securing sales.

Various difficulties are encountered in applying these criteria. The delineation of the boundary of a store trading area to correspond to a fixed drawing-power standard requires much detailed data as well as practical skill in interpreting the effect of numerous geographic and other factors on the extent and shape of a store trading area.

Per capita sales and share of market criteria vary according to the types, sizes, and locations of stores and with the policies and capabilities of each firm. One general merchandise store company considers as within its trading area all areas from which it obtains $1.00 per capita per year or a 0.5 percent share of the market. One supermarket chain considers as within its trading area all areas from which it obtains $5.00 per capita per year or a 1.5 percent share of the market. Should the population within the trading area thus defined increase considerably and should additional competition be established in the process, a decreased per capita sales and/or share of the market might not alter the store's total sales obtained from the trading area, but the definition would no longer be valid. Actually, a thickly populated core will lower the minimum per capita sales necessary to support a store, and a thinly populated core will raise them.

[2] The following are examples:

A. *Drawing power.*
   1. The trading area is "that area from which the community receives approximately 90 percent of its total retail patronage." Isadore V. Fine, *Retail Trade Analysis*, University of Wisconsin, Bureau of Business Research and Service (Madison, 1954), p. 11.
   2. "The area of influence from which a shopping center could expect to derive as much as 85% of its total sales volume is defined here as the trade area of the center." Victor Gruen and Larry Smith, *Shopping Towns U.S.A.* (New York: Reinhold Publishing Co., 1960), p. 278.

B. *Per capita sales.* The trading area is "that area which will provide [a general merchandise store] a minimum annual per capita sale of one dollar." Howard L. Green, Montgomery Ward Co., correspondence with the authors, April, 1959.

C. *Movement.* "Generally speaking . . . a large majority of customers are willing to travel 12–15 minutes, and a maximum of 25 minutes to reach a regional shopping center." Gruen and Smith, *op. cit.*, p. 33.

D. *Population.* "2,000 families would spend $2,000,000 for food annually, or enough to support a large supermarket—if all families were to patronize one store." Max S. Wehrly and J. Ross McKeever, eds., *The Community Builders Handbook*, Community Builders' Council of the Urban Land Institute (Washington, D. C., 1954), p. 134.

E. *Competition.* "The actual trading area is entirely determined by the operation of attraction and resistance of competitive trading areas . . . the trading area is not a permanent geographical fact but is created entirely by the response and behavior of individuals." Richard L. Nelson, *The Selection of Retail Locations*, F. W. Dodge Corp. (New York, 1958), p. 183.

Trading-area criteria based upon movement, such as distance or travel time, might be useful if it were possible to develop a valid formula that would allow for differences in competition, traffic friction,[3] and frequency of customer shopping.

Population can be a useful trading-area criterion if the number is weighted for competition. The trading area of a full-scale department store would require and include 100,000 persons if *all* expenditures of these 100,000 persons for department store goods were made at the store. If this were only one of five equal competing stores, then an area with a population of 500,000 would be needed to constitute the trading area.

Location of competition as a criterion of store trading area is difficult to apply in metropolitan areas, where competition exists not only among stores of approximately equal rank, but also with stores of much greater and lesser rank.

Broad definitions of trading areas suffer from vagueness. Nonetheless, generalizations are useful guides. Some of these definitions stress movement; others, spheres of influence; still others, people.[4] As a broad definition, we suggest that the trading area is the area from which a store gets its business within a given span of time. This does not exclude the reality of overlap. It also emphasizes the "area" in trading area. People must come to a store from a specific area. If other stores offer equal attractions, then the trading area of a given store will be related to this store's distance and convenience of access from the origin and destination of the potential customers.

Because a store's trading area frames the lines of customer movement, the impact of changing shopping-habit patterns upon the channels of this movement is of major importance. Customers today originate the bulk of their shopping trips from home. Their mode of travel, with some rare exceptions, is by automobile, public transportation, or on foot. In the future, they may relate more of their shopping habits to places of work or recreation. There may also be shifts and changes in mode of travel. Regardless of whether the origin of the trip or an intermediate or final destination is the more important, the fact that every shopping trip is linked to a *place* gives a geographical character to the trading area. The mode of travel affects the distance that people will travel readily to shop.

Finally, a generalization of a store trading area must take into account the fluctuations in and somewhat amorphous nature of the boundaries arising from the day-to-day actions of the retailer, his competitors, and the customers. This is discussed more fully later.

FEATURES OF TRADING AREAS

A store trading area is a nodal region (i.e., it contains a focus—the store—which is the node of organization and the surrounding area is tied to the focus by lines of transportation). Its features include: (1) cores, (2) morphology—shape, size, and arteries of movement, (3) competing stores, (4) internal components, and (5) external boundaries.

The Core

The core consists of the point at which the store is located. Characterizing the core are: (a) type of location, (b) size of store, (c) retail structure, and (d) store associations.

Location types. We find the following classification useful in analyzing store trading areas in relation to location types:[5]

[3] Traffic friction results from such disconcerting elements as turning against traffic flow, heavy truck or bus movement, and stops and starts due to street lights, pedestrians, or access lanes.

[4] Examples of such definitions are:

A. *Movement.* "In this study, a market or trading area is thought of as being composed of an area provided with means for the movement of persons and commodities, and occupied by people who have common and daily economic interests." John Paver and Miller McClintock, *Traffic and Trade* (New York: McGraw–Hill, 1935), p. 5.

B. *Sphere of influence.* "A trade center region, i.e., a trade center and its continuous hinterland, is one in which spheres of influence should be exerted by the former upon the latter." Eugene Van Cleef, *Trade Centers and Trade Routes* (New York: Appleton–Century Co., 1937), p. 38.

C. *People.* "A store does not obtain business from any area. It obtains its business from people." Reavis Cox, correspondence, December, 1959.

[5] This is not a quantified classification; therefore, it lacks precision of definition. However, it is a classification which can be readily applied. For an approach to a quantified, functional classification, see *Shopping Center Report* (preliminary report), Cleveland Regional Planning Commission, Cleveland, Ohio, 1958, p. 26.

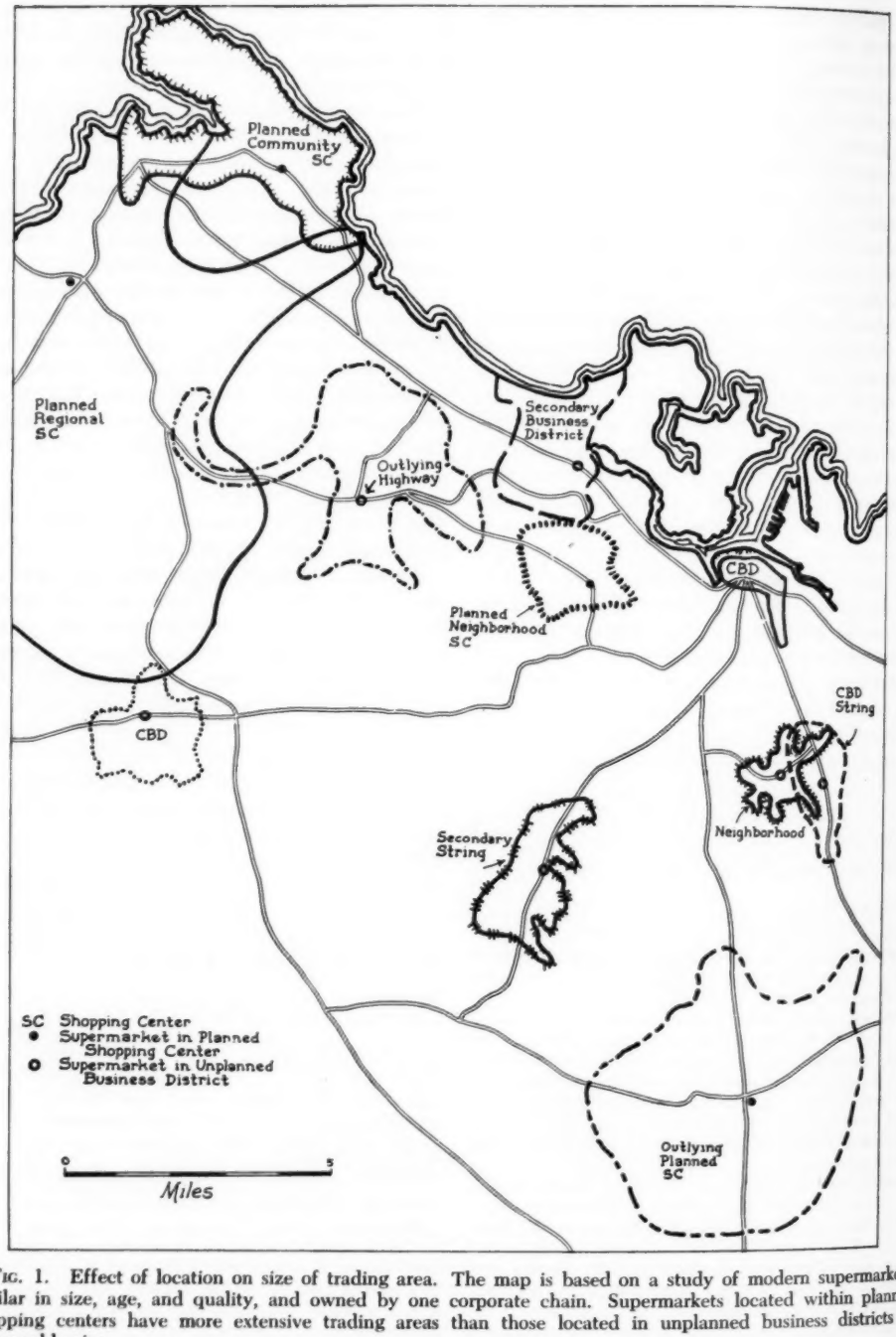


FIG. 1. Effect of location on size of trading area. The map is based on a study of modern supermarkets, similar in size, age, and quality, and owned by one corporate chain. Supermarkets located within planned shopping centers have more extensive trading areas than those located in unplanned business districts of comparable size.

A. Unplanned business districts
   1. Central business district
   2. CBD string stores. Adjoin the CBD
   3. Secondary business district. Serves portions of a central city, or a suburb
   4. Secondary string stores. Adjoin secondary business districts
   5. Neighborhood stores. Occur in small clusters or in isolation
   6. Outlying highway stores. Occur in strings or in isolation

B. Planned shopping centers
   1. CBD planned shopping center. Arises through urban renewal
   2. Regional planned shopping center. In strong competition with the CBD
   3. Community planned shopping center. In competition mainly with secondary business districts, or with the CBD in smaller cities
   4. Neighborhood planned shopping center. Frequently called neighborhood "strip"
   5. Outlying planned shopping center. Draws, in part, upon the passing parade of highway traffic

In towns and small cities where the CBD is generally the only important business district, the size and other characteristics of a store are more important than type of location, since, in order to obtain a significant amount of trade, a store must draw its business from the same trading area as the CBD. Therefore, the location types presented in this classification, except for the CBD, are of practical significance only in metropolitan areas and larger cities.

Figure 1 presents examples of store trading areas in these types of locations, based on a number of supermarkets, similar in size, age, and quality and located within a relatively evenly populated metropolitan area.

Store size. Store size has a bearing upon trading area insofar as size of facility reflects the amount and variety of goods and services that can be provided for customers. However, trading-area extent does not increase proportionately with increasing store size.

A supermarket with 8,000 square feet of sales area can carry a selection of items to compete in trading-area extent with a store very substantially larger in size. A supermarket of 25,000 square feet of sales area, however, has a significantly larger trading area. A discount house with 40,000 square feet can command as extensive a trading area as one 50 percent larger. Smaller discount houses, however, have smaller trading areas. Branch department stores may range in selling area from 60,000 to 100,000 square feet without appreciable change in their trading-area extent. A full-line department store of 250,000 square feet creates a far different trading area.

Retail structure. A store's trading area is affected by (1) the retail–gravitation power of the city in which the store is located and (2) its proximity to other stores.

The retail–gravitation power of a city derives from the number and type of stores it has. This number and size is usually related to the size and function of the city. Large cities have more extensive trading hinterlands than small cities. So do cities whose factories reach out to distant towns and villages for additional labor forces. When factories extend their employment base rapidly and as retailers recognize and take advantage of the changes in market potential, store trading areas are extended.

In general, the larger the population of a city or metropolitan area, the more extensive is the trading area of the CBD. Similarly, the larger the suburbs in metropolitan areas, the more extensive are the trading areas of their stores dealing in shopping goods and even convenience goods.

The decentralization of retail facilities via large planned shopping centers in suburban areas is increasing the extent of metropolitan areas into rurban and rural hinterlands.

As the size of a city increases, the share of the available retail business secured by stores located outside the CBD increases also. Larger cities present more traffic friction to shoppers in getting to and from the CBD. This traffic friction creates greater opportunity for suburban stores. As cities grow, therefore, the CBD stores receive a smaller share of the retail business of suburban and rural portions of the trading area.

Table 1 presents data on the trading areas of department stores in 13 cities of different size. These data show an inverse relationship between the drawing power of a store and the extent of its trading area. When these statistics are weighted by the proximity of competing cities and by traffic friction outside

TABLE 1.—DEPARTMENT STORE DRAWING POWER, BY SIZE OF CITY AND EXTENT OF TRADING AREA

Size of city (population)	Trading area extent (miles)	Percent of department store sales	
		Outside city	In city
15,000	15	45	55
40,000	20	30	70
50,000	40	35	65
60,000	15	40	60
60,000	20	50	50
150,000	25	30	70
200,000	20	35	65
225,000	15	25	75
350,000	15	10	90
350,000	50	20	80
450,000	50	25	75
750,000	60	10	90
1,000,000	50	10	90

the city, the inverse relationship between drawing power and trading-area extent is more pronounced than indicated by the table.

A study of supermarkets in or near the CBD's of 40 towns and small cities affirms that as the size of a city increases, store drawing power beyond the city decreases, even though trading-area extent increases with city size (see Table 2). Because of greater frequency of trips to supermarkets than to department stores, traffic friction and competitive opportunities play a stronger role in affecting the drawing power of supermarkets than of department stores.

As stated previously, increased size of metropolitan areas, as well as of cities, increases the extent of their CBD store trading areas. However, increased size of metropolitan area does not appear to result in decreased drawing power beyond the metropolitan area. An indication of this comes from an examination of the retail sales of standard metropolitan areas in relation to their populations. If drawing power to the hinterland increased with a decrease in size of metropolitan area, then per capita sales—calculated by dividing total retail sales of all stores in the metropolitan area by the population residing within the metropolitan area—would increase accordingly. However, this is not the case.

An analysis of the 1954 *U. S. Census of Business* data of annual per capita retail sales by metropolitan areas in the United States indicates a wide range, with no correlation with increased size. Among the first 40 Standard Metropolitan Areas, the range of annual per

TABLE 2.—DRAWING POWER OF SUPERMARKETS, BY SIZE OF CITIES

Size of city (population)	Supermarket sales (percent)	
	Outside city	In city
Under 2,500	50	50
2,500 to 10,000	40	60
10,000 to 20,000	35	65
20,000 to 40,000	25	75
40,000 to 60,000	15	85
Over 60,000	10	90

capita retail sales is from $900 to $1,500. The figure for the largest Standard Metropolitan Area, New York, is $1,150; for the 40th, Fort Worth, it is $1,270. The median is $1,230. There appears to be no discernible order in the figures in relation to size of metropolitan areas.

While these data are inadequate for more than a tentative conclusion, the authors conjecture that highest per capita sales for the metropolitan area as a whole occur in metropolitan areas which have the largest number of planned regional shopping centers.[6] The occurrence of such shopping centers is not directly related to the size of metropolitan areas. Most of these planned regional shopping centers are on the fringes of metropolitan areas. They have extended their retailing influence much farther than ever before. Because most of these planned regional shopping centers are still surrounded by relatively sparse populations, and because their strongest pull is away from the CBD, their trading areas are extensive and their drawing power to the hinterland is high.

In the long run, as population density builds up around these planned regional shopping centers, drawing power can be anticipated to increase close to the center and to decrease in the hinterlands.

Store associations. When a site is occupied by a combination of stores, the trading area of each store is affected by the trading area of its neighbors. In theory, the most extensive boundary of the most powerful store will also serve as the boundary of all other stores. In practice, however, the percentage of business that many stores obtain from the outermost

[6] Verification of this conjecture will be possible by studying the latest available data on planned shopping centers, the 1958 *Census of Business* data by metropolitan areas, and the 1960 population data, when published.

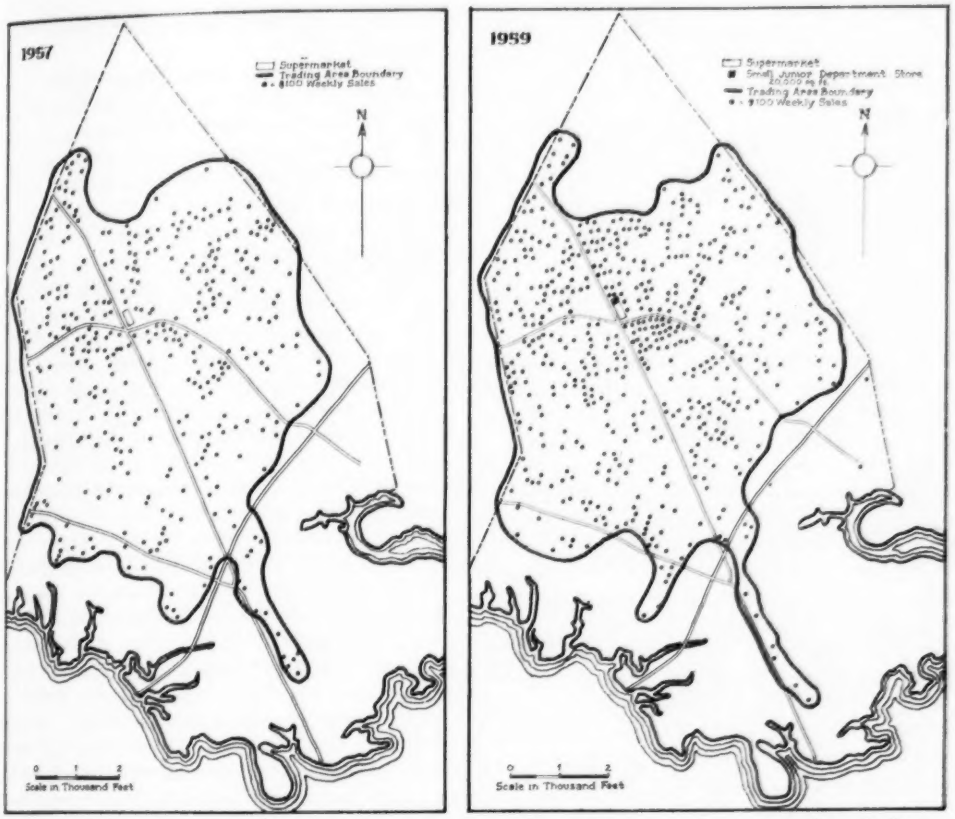


FIG. 2. Effect of a small junior department store upon the trading area of a supermarket. Built as a free-standing unit in 1957, the supermarket shown on the map was joined in 1960 by a small junior department store. While the supermarket's trading area boundary has remained unchanged, per capita sales have increased by 15 percent.

portion of a retail district's trading area tends to be only slightly greater than if these stores were free-standing. The association of one store with another either helps to extend its drawing power or increases its per capita sales, or both. Where the drawing power of two stores is different, there is no assurance that the more powerful store will significantly affect the drawing power of the weaker store. Differences in customer shopping habits in patronizing the various stores concerned, as well as the location of stores that are competitive to the less powerful store, must be taken into account.[7]

[7] For a more detailed discussion, see Saul B. Cohen and William Applebaum, "Evaluating Store Sites and Determining Store Rents," *Economic Geography*, January, 1960, pp. 25–30.

A study of the trading areas of 32 modern supermarkets, half in planned shopping centers containing department stores or large variety goods stores and half on free-standing highway sites, indicates that associated stores with considerably greater trading-area extent do expand the trading-area boundaries of the supermarkets; in many cases, a mile or so more than normal. Furthermore, by this association the drawing power of the supermarkets in the large planned shopping centers was 15 percent from beyond 3 miles, compared with 10 percent for the supermarkets on the free-standing sites. On the other hand, while small junior department stores may contribute to an increase in a supermarket's per capita sales they have little impact upon a supermarket's trading area (see Fig. 2).

TABLE 3.—EFFECT OF POPULATION DENSITY ON TRADING AREA EXTENT AND DRAWING POWER OF NEIGHBORHOOD-TYPE SUPERMARKETS

Number of stores	Population within one mile of store	Trading area extent (miles)	Supermarket sales (percent)	
			Beyond one mile	Within one mile
25	7,000	1½	75	25
20	17,000	1¾	60	40
20	27,000	2	50	50

Trading-Area Morphology

The size and shape of store trading areas are determined by location, population density, accessibility, natural and cultural features, competition, and store "personality."

Size of trading area. As discussed in the sections devoted to store size and retail structure, the size of the trading area of a store (or of a group of stores) increases with the size of the store and with the size of the urbanized area, and the percentage of sales obtained from the more distant portions of the trading area varies inversely with the size of the city. In relating the latter phenomenon to specific stores, we find that population density affects a store's drawing power. This is demonstrated by the experiences of 65 neighborhood-type supermarkets, as presented in Table 3.

Dense populations attract larger and greater numbers of the same type of store to one focal point. The effect of such multiplication of store facilities is to extend the trading area for each store. Traffic friction, on the other hand, acts as a counterforce and serves to reduce the share of business that is obtained from more distant parts of this trading area.

Another observation that can be made is that the higher the income of the tributary population, the more extensive is the store's trading area and the greater the drawing power to the more distant parts of the trading area. Also, the more densely populated the rural parts of the trading area, the greater the drawing power of a store.

The role that accessibility plays in determining a store's trading area can be seen from mapping the trading areas of stores that lie close to major intersections. Thus, differences in accessibility for two shopping strips on opposite sides of a rotary helped to account for a difference in trading-area extent of one and one-half miles.[8]

Shape of trading area. Theoretically, the hexagonal, triangular, or square trading areas offer the best possibilities for covering a market. In practice, however, street patterns play the most significant role in moulding such shapes, and frequently they run counter to what would be theoretically desirable. In cities with grid-type street patterns, store trading areas tend to be rectangular, with amoeba-like extensions, and adequate coverage with least overlap is obtained. Natural and cultural features, such as rivers, railroads, greenbelts, escarpments, unpaved roads, and differences in population characteristics, can abort the shape of store trading areas at distances that vary with location types. Numerous examples can be cited. Caution should be exercised in evaluating such barriers. Where rivers are crossed by wide, toll-free bridges, where railroad tracks are in use at infrequent intervals, where well-maintained highways cut escarpments, there the barrier qualities of such features are nullified. Certain cultural features are more lasting barriers than physical features. For example, a political and cultural break between a fashionable suburb and a slum area tends to form a sharper trading area boundary than does a river. Also, such trading area boundaries occur in border zones between the white and Negro populaces of large northern U. S. cities. On the other hand, the general rise in living standards acts as a counterweight in blurring these distinctions.

The effects of competition upon the shapes of store trading areas can be seen most clearly from the relation of the trading areas of suburban stores to the CBD. Usually, the trading-area shapes of suburban stores are elliptical, with the longer axis trending away from the CBD. The trading-area shapes of suburban CBD's are similarly affected by the central-city CBD (see Figs. 3 and 4). In some instances, the competition of the CBD may create a hole within the shape of a large suburban store's trading area. This occurs when the CBD is in competition with large suburban department stores (see Fig. 5).

Store "personality" refers to the customer's image of a store, formed on the basis of merchandise, price, services, and physical facili-

[8] *Ibid.*, p. 29.

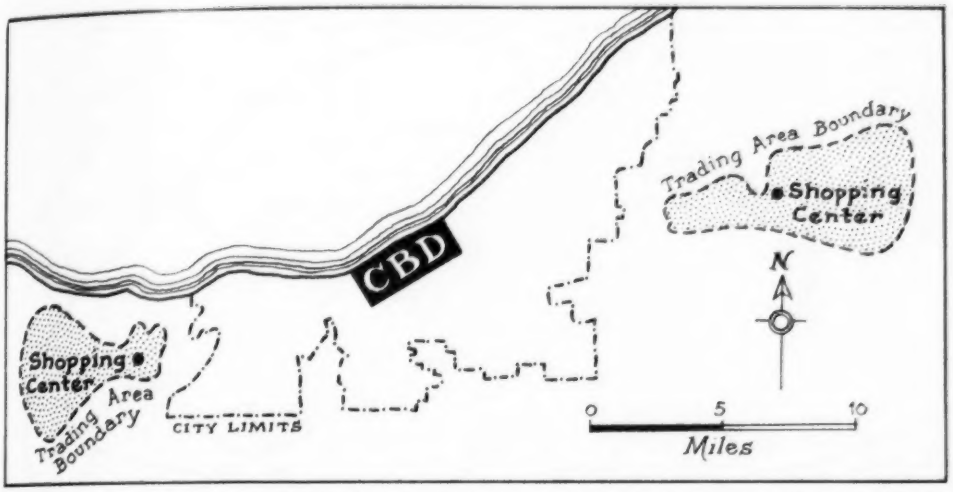


FIG. 3. Shape of trading areas of suburban stores. Trading areas of suburban stores tend to be elliptical, with the longer axis trending away from the CBD. Customer-spotting maps of two medium-size suburban planned shopping centers are the basis for this figure.

ties. Store "personality" affects the trading area, even though this effect eludes measurement. Marked uniqueness of store personality is likely to extend the trading area significantly. This is the case, for example, where a mill discount outlet carries the same broad lines of merchandise as a junior department store of the same size. The former features lower prices, frequently on lower quality merchandise.

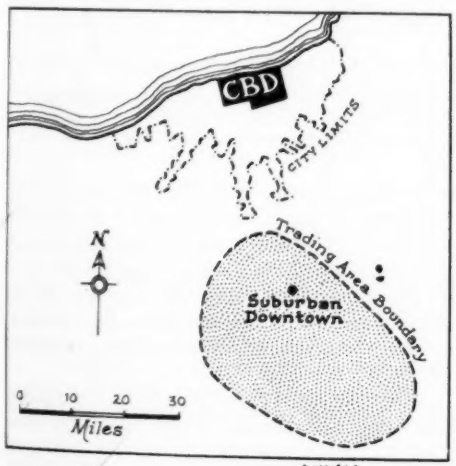


FIG. 4. Impact of the CBD of a large city upon the shape of a suburban downtown trading area. This elliptical trading area presents the trading area of a downtown in a suburb of 60,000 population; the suburb is located within a large metropolitan area.

While it is possible to isolate the specific effects of population, accessibility, physical and cultural features, and competition, usually a combination of these elements is involved. This is demonstrated through the mapping of a planned shopping center of medium size (see Fig. 6).

Arteries of movement. In general, store trading areas are elongated in the direction of customer movement (see Fig. 1). This tendency is qualified, however, by the distribution of population and competition, by access to and traffic friction along the artery.[9] If the population tributary to a major highway is sparse, then, despite heavy traffic flow along the highway, the store does not draw much trade from this traffic flow.

Competing Stores

The effect of competing stores upon each other's trading area should be considered in a time dimension. The initial trading area of a store will be related to the location of existing competing facilities. A store's trading-area boundaries lie within a zone that forms the

[9] For examples of the role played by access and traffic conditions, see *ibid.*, pp. 1–8.

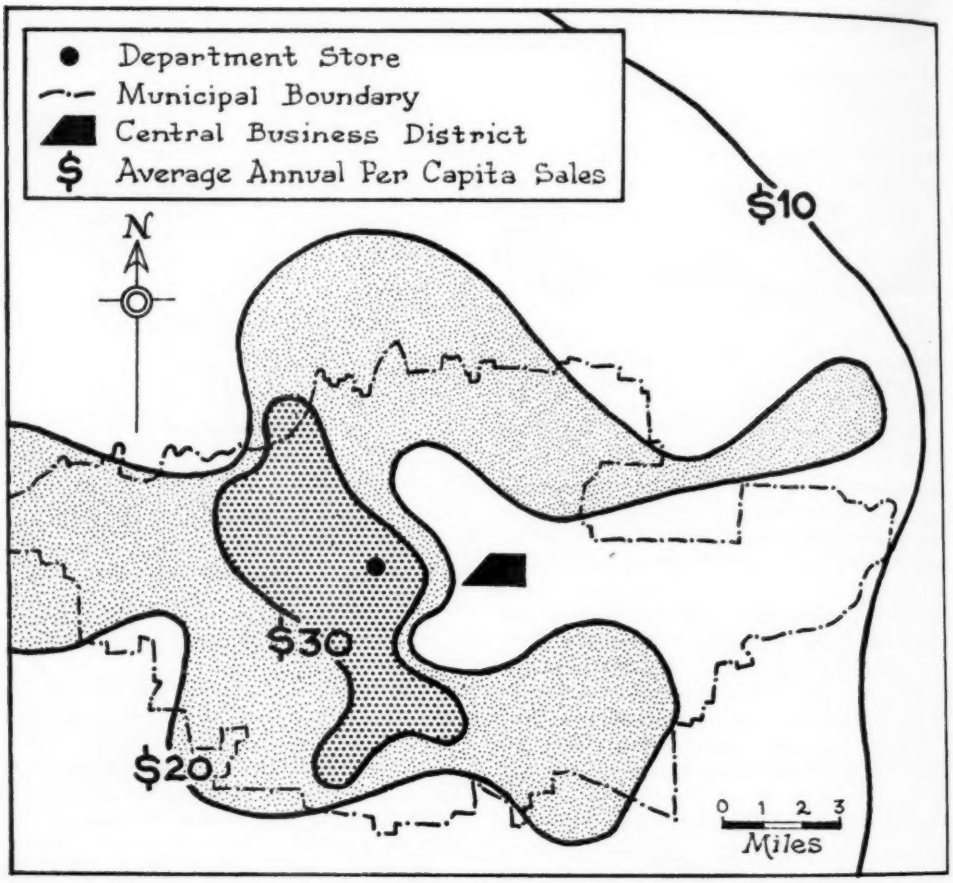


FIG. 5. Effect of CBD as competitor of a large suburban department store illustrated by per capita sales contour map. Per capita sales contours are "pinched in" by the CBD; only $10 to $20 in annual per capita sales are being realized by the department store from the areas surrounding and east of the CBD.

"trading divide" between this store and competing stores. Only when there is little or no population between two competing stores, or when the stores are very close together, do we find that competing stores exercise no impact upon each other's trading-area boundaries. Our studies of 15 large urbanized areas show that supermarkets which are within a half mile of one another serve the same general trading area and do not influence significantly each other's trading-area boundaries. The same applies to department stores within five miles of each other.

Once the trading-area boundaries of a store have been established, then these boundaries are not affected greatly by additional nearby competing stores. However, the store's market penetration is very likely to be affected adversely. This is not to say that new, nearby competing facilities do not abort, bend in, or create a gap within certain portions of the trading area of an existing store. But in broad, meaningful terms, the boundaries remain fairly constant (see Fig. 7).

What has been said above refers to competing stores that belong to different companies. The competitive impact of a store belonging to the same company is quite different. A store's strongest competitor is a sister store with similar facilities. The trading-area boundaries of a store that is confronted with competition by a new store of the same company

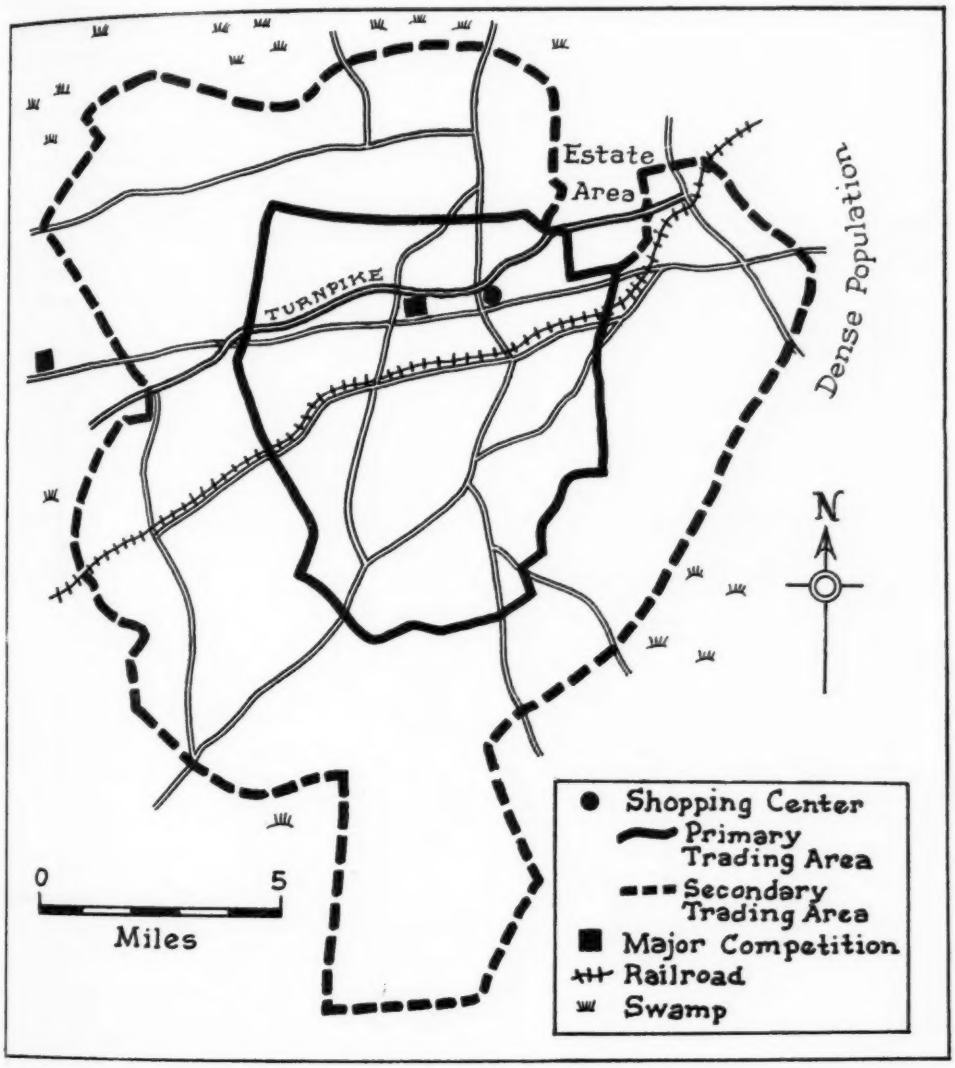


FIG. 6. Effect of population, accessibility, physical and cultural features, and competition on shape of store trading area. The trading area shown is that of an outlying highway planned shopping center with moderately priced goods. Swamps, dense population to the east, and style-conscious estate dwellers to the northeast abort the secondary trading area. Major competition two miles to the west is too near to affect the boundary of the primary trading area. However, the location of a major shopping area 12 miles to the west does affect the western boundary of the planned shopping center's primary trading area—the boundary follows the breaking point between the two business centers.

tend to shift and seek a trading divide. Figure 8 shows how the trading-area boundaries of a CBD supermarket changed as a second and then a third supermarket were built in the town by the same company.

Internal Components

A store trading area contains within it individual houses, city blocks, neighborhoods, large urban areas, a metropolitan region, and sometimes even larger geographic units. These

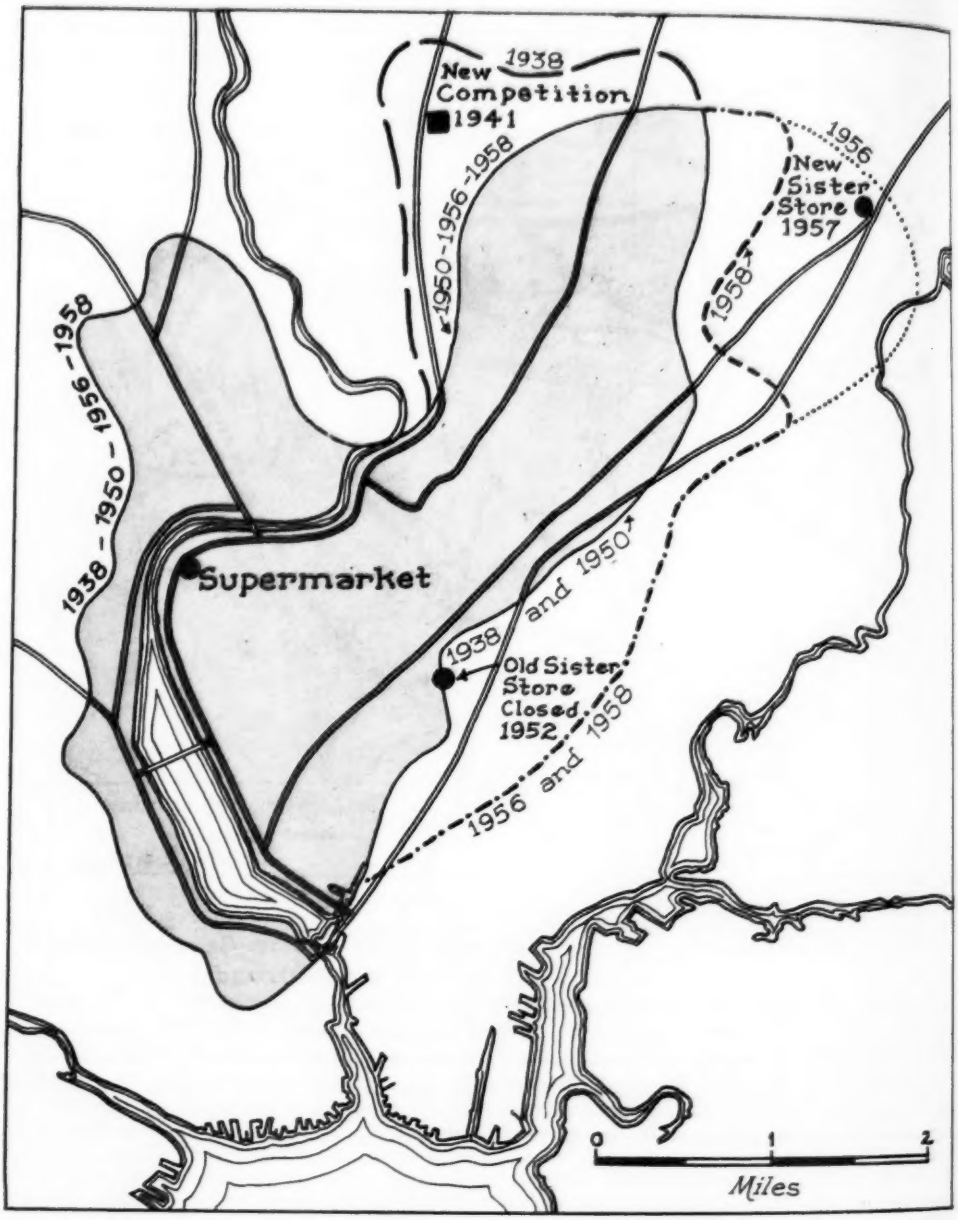


FIG. 7. Stability of trading area. A store's trading area boundary remains fairly constant over long periods of time. Customer-spotting studies in 1938, 1950, 1956, and 1958 are the basis for this map. The supermarket was built in 1938. Significant boundary changes have occurred only in the areas affected by the 1957 opening and the 1952 closing of sister stores, and by the construction in 1941 of a considerably larger competitive store with better facilities.

parts of a store's trading area are the internal components. They have different sales values to a store and must be studied to understand the complex nature of a store's trading area.

To differentiate these different components, various criteria are employed: share of market, driving time, traffic flow, location of competition, and drawing power.[10]

Major share of market is often used to mark off one part of the trading area from another. While a retail district may obtain more than half of the available business of a given area, individual stores rarely secure a major share of the available business of any substantial part of the trading area.

Driving time (determined by mode of travel, traffic friction, individual driving habits, weather, accessibility, and parking) can be a useful criterion if the values of the units of time used as measurements are related to actual sales experience.

Breaks in traffic-flow density can be useful as a rough guide in differentiating trading-area components; however, total traffic density may not be a good indicator of traffic density for shopping purposes.

Location of competition as reference points in establishing the components of a store's trading area is useful only for broad generalizations which do not give due consideration to the element of overlap of store trading areas.

Drawing power is a useful criterion in differentiating trading-area components. However, its usefulness depends on giving proper weights to differences in population density and purchasing power.

Various terms are used to describe trading-area components. *Primary*, *secondary*, and *fringe* are perhaps the most common. These terms are often used ambiguously. For a large retail district, say a CBD, "primary" may be used to designate that part of the trading area from which a major share of the total market potential is obtained. For individual stores, "primary" usually means the largest single share of only a fraction of the total market potential.

We suggest that the components of the trading area be differentiated on the basis of the sales penetration and drawing power of a store (or group of stores) within the trading area. This combination of judiciously selected measurements offers a sound, practical method for evaluating and mapping the internal components.

The per capita sales of a store, i.e., a store's retail sales per person (per week or per annum), measure the extent of a store's sales penetration of a trading area. This measurement represents differences in the store's influence upon customers in different parts of the trading area. Per capita sales can be weighted for differences in purchasing power for more meaningful comparisons, and can be adjusted for price index changes, where com-

---

[10] Examples of various criteria for differentiating the internal components of a store's trading area are:

A. *Share of the market.* "The uncontested trading area is that part of the trading area in which the great majority of business goes to the retail center." August Lösch, *The Economics of Location*, transl. by William H. Woglom (New Haven: Yale University Press, 1954), map p. 418.

B. *Driving time.* "The critical point at which the maximum effectiveness [of the shopping center] decreases varies less than might be expected in terms of driving distance and driving time for regional shopping centers regardless of size. The primary trade area is the area of strongest influence, normally three to six miles [12 to 15 minutes] from the site." Gruen and Smith, *op. cit.*, pp. 33 and 278.

C. *Traffic flow.* "The Traffic Stabilization boundary is that portion of the Urban Market that lies at the center of the Trading Area in which there is from the center of the city . . . a relatively high and continuous density of population. This can be marked by the traffic stabilization point . . . the point outward from which there is no further rapid decrease of traffic volume." Paver and McClintock, *op. cit.*, p. 52.

D. *Drawing power.* "Normally, the trade area is divided into zones which indicate variations in the drawing power and strength of the development in different portions of its tributary areas. These zones are usually called the primary, secondary and tertiary zones of the trade area." *Definitions of Terms* (Washington, D. C.: Larry Smith & Co., 1960), p. 2.

E. *Location of competition.* "The primary trade area—the close, walk-in area and the area that has daily convenience stores closer than a site under investigation; the secondary trade area—the area which has . . . no important soft line and hard line stores or shopping goods stores with more convenient access than the site; the country or fringe trade area—the area which may be closer to department and apparel stores, but from which customers might be drawn because of greater parking conveniences and better merchandise to be offered." Wehrly and McKeever, *op. cit.*, p. 136.

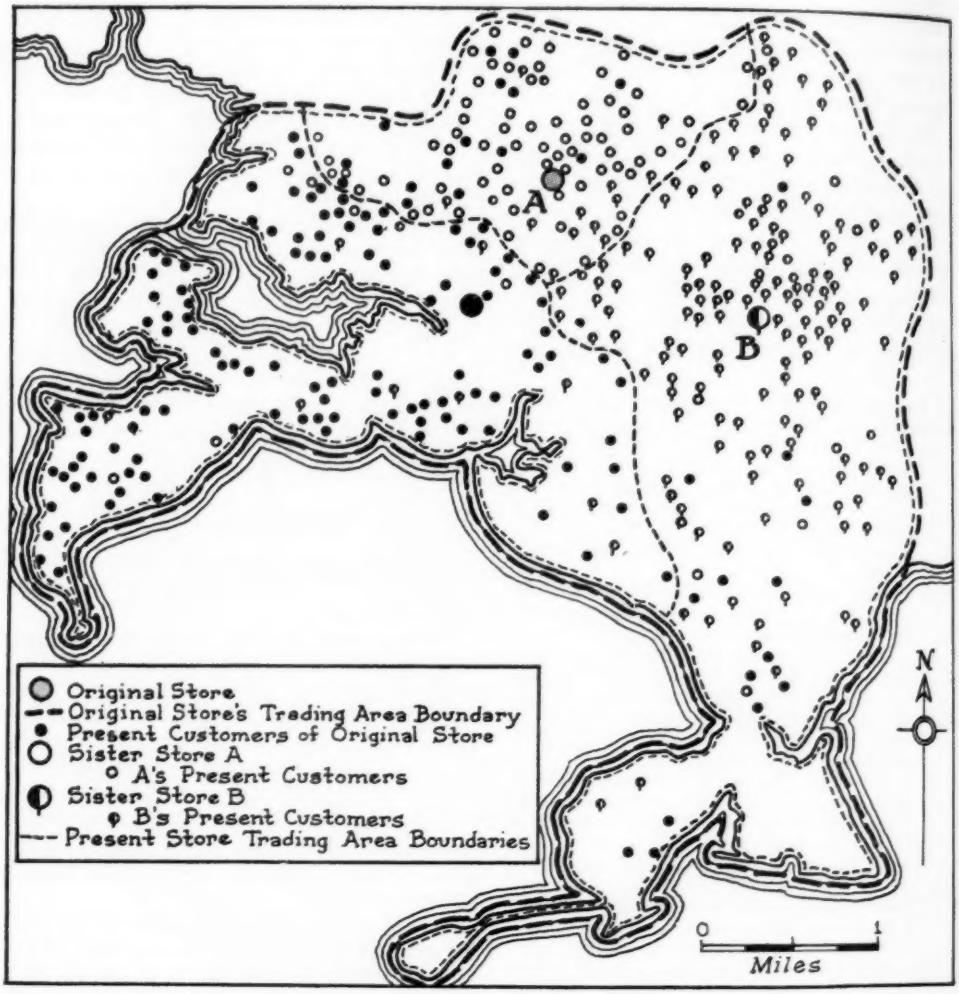


FIG. 8. Impact of competition on a store's trading area. A store's strongest competitors are its sister stores. With the building of sister stores A and B, the trading area boundary of the original supermarket shown has shrunk to one-fourth of its original extent.

parisons over time are needed. Such weightings and adjustments are essential to arrive at sales penetration in terms of share of market. Drawing power quantifies the relative sales value of that component to the store. Thus the selected drawing-power criterion designates the proportionate value of the component, and the selected per capita sales contour outlines the boundary of the component (see Fig. 9).

Customer-spotting maps, which show the varying density of customer distribution, can serve as the guides to the selection of drawing-power criteria for establishing internal components of a trading area. Customer-distribution density varies for different types of stores and for different locations. Our experience leads us to designate the primary trading area of supermarkets in large planned shopping centers as that area which contains somewhat more than half of all the customers; for neighborhood supermarkets the primary trading area is designated to contain up to 85

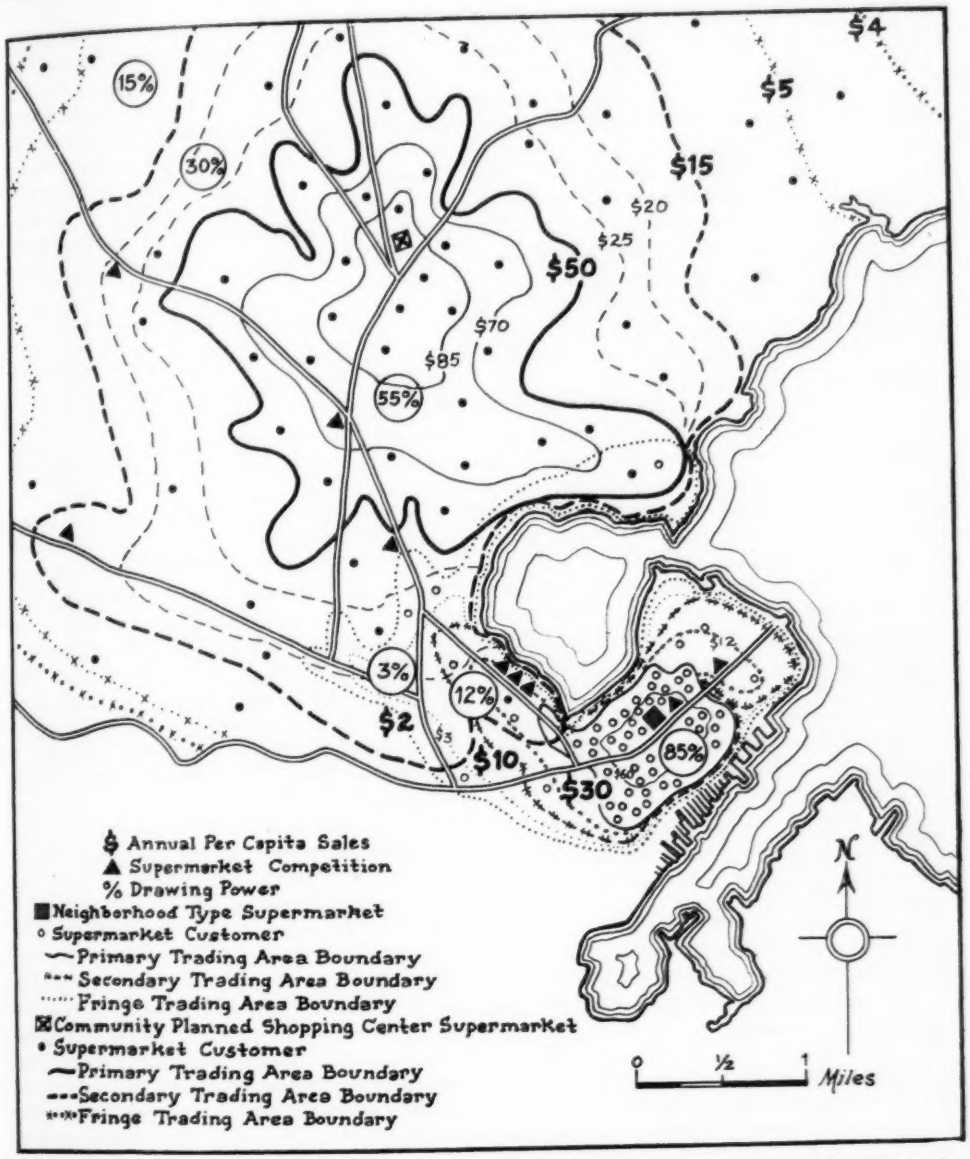


FIG. 9. Sales penetration and drawing power of two supermarkets, one in a planned shopping center, the other in an older neighborhood business district. Per capita sales of the neighborhood store are lower and drawing power is more intensive, because of the greater competition and traffic friction encountered in the neighborhood location. The store in the planned shopping center, on the other hand, encounters a minimum of competition and traffic friction.

percent of all the customers. For department stores and large planned shopping centers, we would designate the primary trading area as that component which embraces 50 to 70 percent of all customers; the secondary trading area, 30 to 20 percent; the fringe, 20 to 10 percent.

In mapping store trading areas, we have

found it practical to include in the primary trading area the area in which per capita sales are at least twice those of the secondary trading area. Similarly, the per capita sales in the secondary trading area are at least three times those of the fringe trading area, and, usually, much more. If customers are so thinly diffused in the fringe trading area that meaningful measurement of per capita sales cannot be made, then the fringe trading area is described only in terms of drawing power—i.e., percent of sales obtained from beyond the secondary trading area. The inclusion of the fringe-area sales in the total is essential to bring the measurements up to 100 percent.

Per capita sales vary by type of store and type of location. Therefore, in differentiating between primary, secondary, and fringe trading areas, it is necessary to select different per capita sales as units of measurement for each store type and location type. In general, greatest sales penetration can be expected for those types of stores which are most specialized and least numerous, with high-priced and slow-turnover goods. In the fringe trading areas, however, per capita sales as measures of penetration are likely to be more nearly equal for various types of stores. Studies of fringe areas of both supermarkets and department stores indicate that per capita sales account for 1 to 2 percent of the market.[11] This fact is probably due to relatively less frequent customer visits from the fringe trading areas to department stores than from the fringe trading area to supermarkets.

The boundary between primary trading area and secondary trading area appears to be affected by the location of competition, but not the boundary between the secondary and fringe trading areas.

We have found that for supermarkets in neighborhood locations, the boundary lines of the supermarket's primary trading area were unaffected by the location of competition within a half mile of the store. Where competition was located from three fourths of a mile to one and one-half miles away, the location of these competing stores marked the edge of the primary trading-area boundary. Where competition was located beyond one and one-half miles, the boundary of the primary trading area lay midway between the supermarket and the competing store. For supermarkets located within large planned shopping centers, the location of competing stores within one and one-half to two miles (along the routes of major highways) did not affect the boundary of the former's primary trading areas.

[11] This share of the market is obtained by relating a store's actual per capita sales to estimated available business. Available department store and supermarket business is estimated on the basis of $100 per capita per annum for the former and $325 for the latter.

External Boundaries

Store trading-area boundaries tend to remain fixed. Changes occur mainly because of the impact of competition from new stores belonging to the same company. A radical change in transportation can also change the boundary.

By fixed boundaries, however, is not meant complete rigidity. Store trading-area boundaries shift within a zone. In certain instances, these shifts take place over a relatively brief period. Customer-shopping habits in the United States are such that persons living a considerable distance from supermarkets tend to concentrate their shopping towards the end of the week, while those living close by spread their more frequent shopping trips throughout the week. Thus, there is a measurable difference in the trading area of a supermarket at the beginning versus the end of the week.

Store trading areas also vary acording to changes in seasonal population. A store that serves an area with a large influx of summer residents experiences an "extended" trading area in serving these people. This does not necessarily mean a blanket extension of the trading area. While adding close-by summer resort sections, the influx of seasonal population, increased traffic friction, and more (seasonal) competition at other points may exclude certain distant parts of the store's winter trading area. (See Fig. 10 for examples of seasonal fluctuations in a supermarket's drawing power.) A special events sale, with extraordinary advertising power, also may extend a store's trading area, albeit for a very brief time.

The permanence of a store's trading area lasts only so long as the store carries out the same functions. If, for example, a store is replaced by the same type of store with improved merchandising facilities only, then the

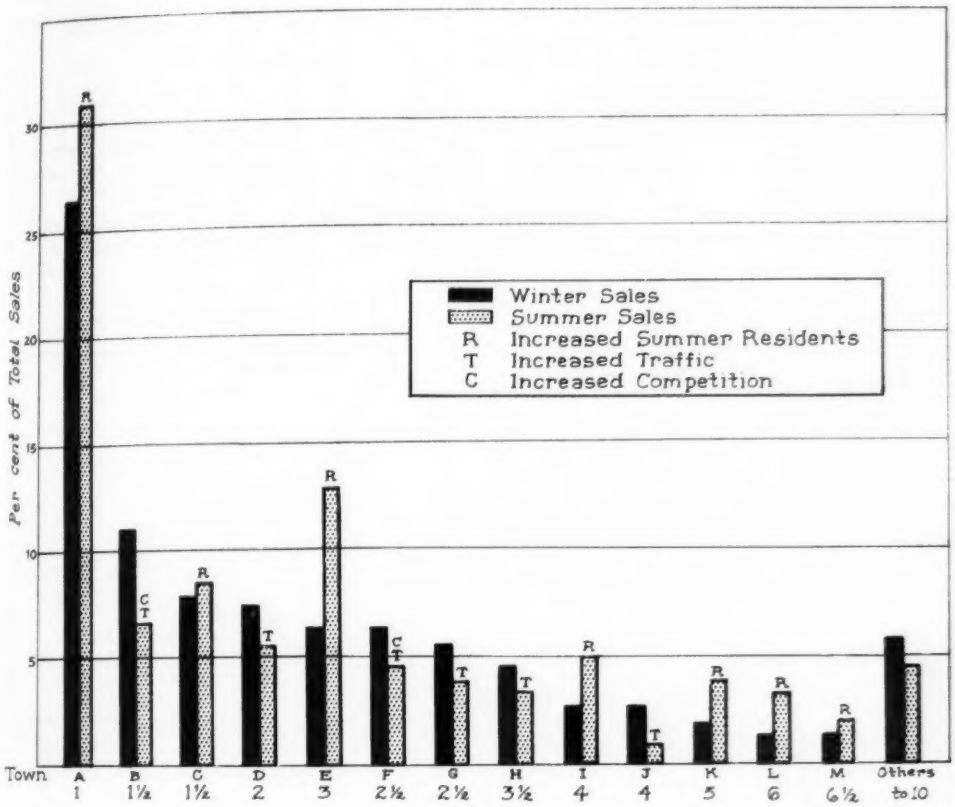


FIG. 10. Seasonal fluctuations in a store's drawing power as influenced by competition and traffic friction. This large supermarket's sales are twice as high in summer as in winter. In general, the drawing power of this store is more extensive during the summer because of an increase in population in the outlying areas. However, competition and traffic friction decrease the drawing power in much of the $1\frac{1}{2}$-to-$3\frac{1}{2}$-mile zones.

trading area may remain unchanged, although the market penetration may improve. This point is illustrated by Figure 7. The original store was one of the early self-service supermarkets located on a highway. When replaced a decade later by a new, more modern unit, its trading area did not appreciably change, but its penetration did change very markedly.

If the functions of a store change significantly, as reflected by changes in the merchandise mix and in service, trading area boundaries tend to change. Figure 11 illustrates this point by comparing the boundary of a CBD service food store that was replaced by a modern supermarket. Similarly, where a modern supermarket was replaced by a very large, combination supermarket–general merchandise store, a marked change in trading area occurred (see Fig. 12, stage 3).

HIERARCHICAL NETWORKS OF TRADING AREAS

Within a given market, each store acts as the focus for a particular trading area. The pattern of the network formed by the trading areas of all stores serving a market is intricate, both because of the overlap of trading areas of stores of the same type and because of the superposition of different orderings of trading areas.

Two forces in particular influence the pattern of store trading-area networks. One force is the central-place role of each store. Lösch and others have stated that market areas for the same branches of business resemble even-

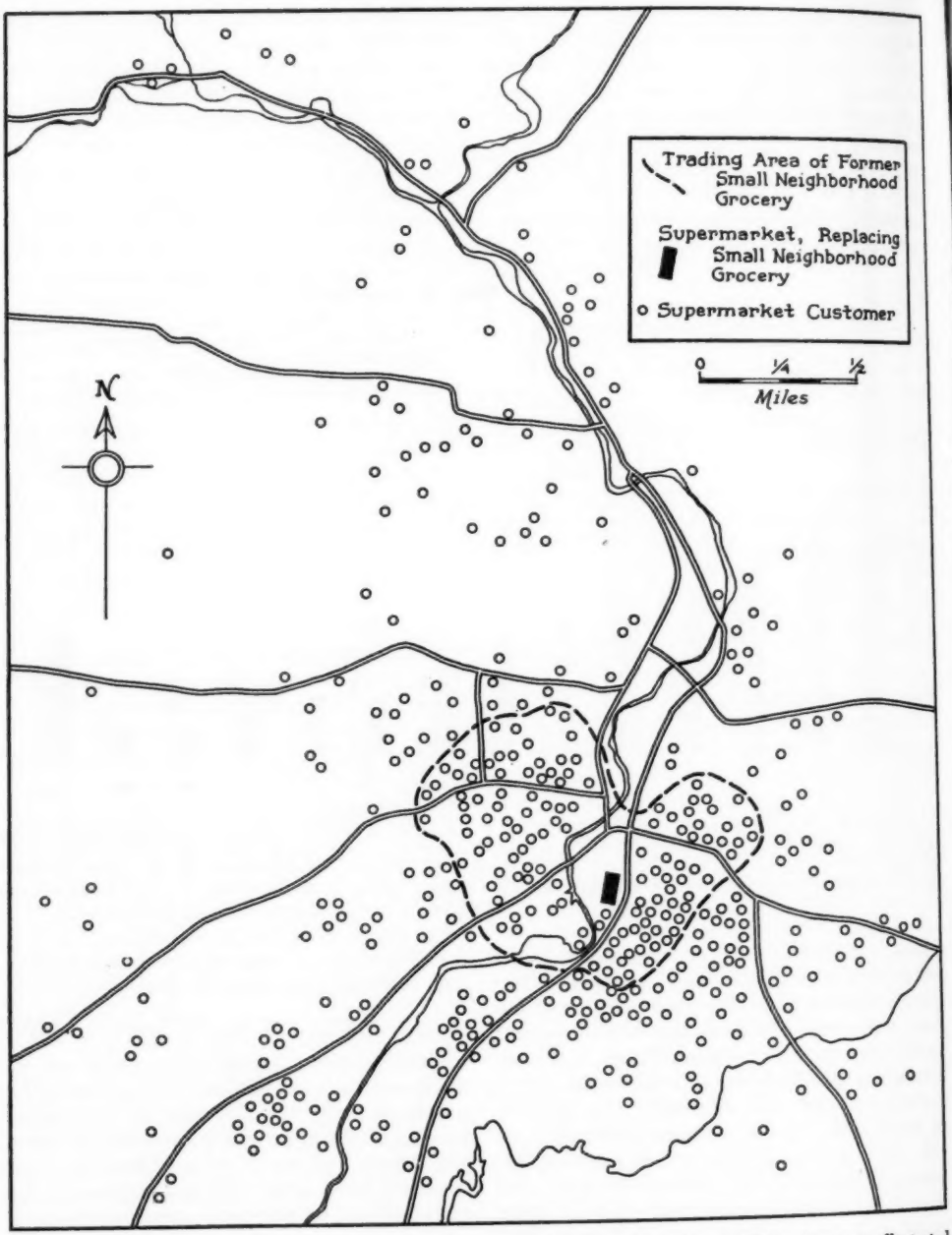


FIG. 11. Effect of change in functions of a store on trading area boundaries. This change is illustrated by the replacement of a small grocery store by a modern supermarket.

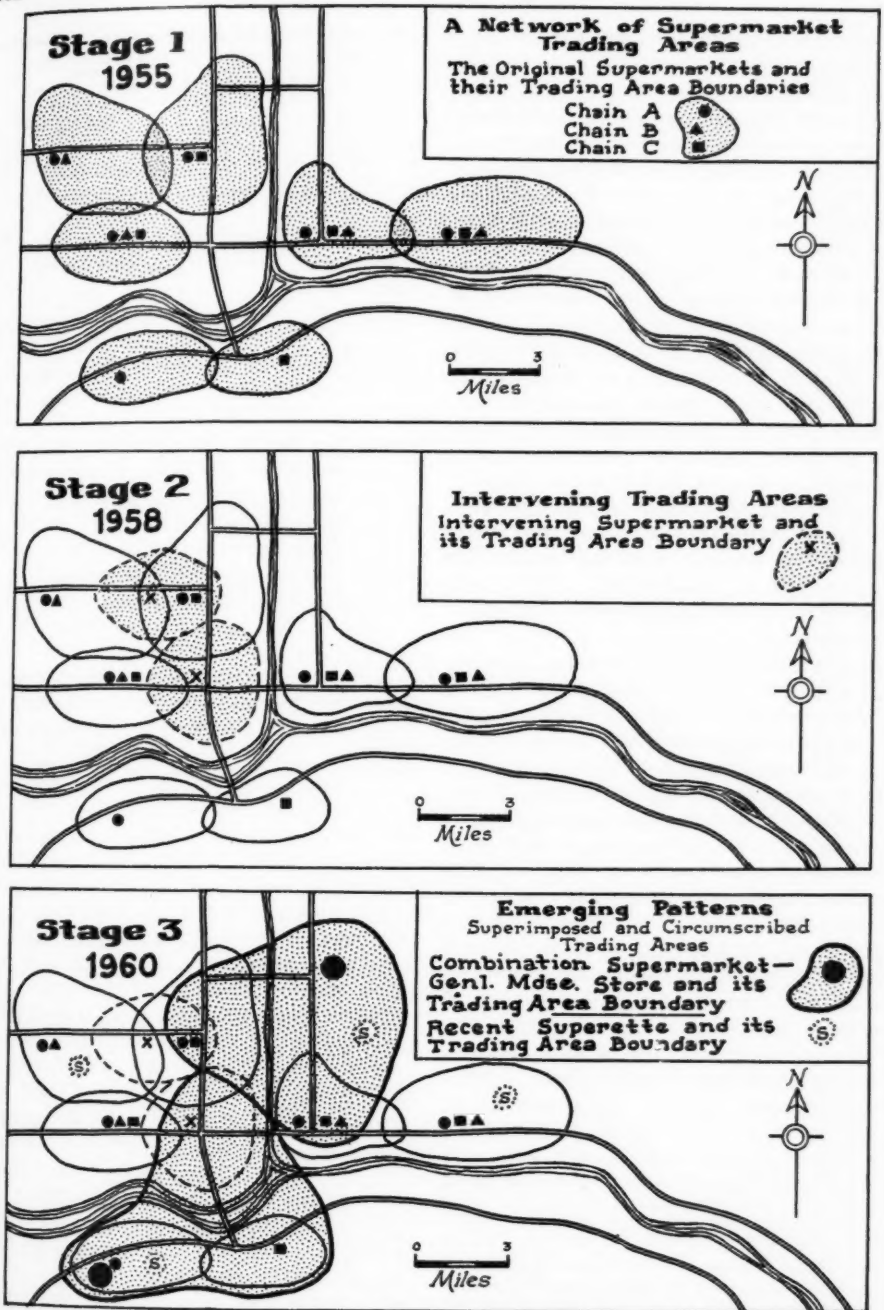


FIG. 12. Rearrangement of store trading area networks as additional stores enter market area. Stage 1 represents the trading area networks formed by groups of supermarkets in a small metropolitan area in 1955. Stage 2 shows that in 1958 new supermarkets found room within the market area by taking up intermediate locations. Stage 3 shows that in 1960 new stores were opened, designed to perform radically different functions, reflected by their superimposed or circumscribed trading areas.

meshed nets of hexagons, and that stores serving these networks tend to be as close to their population centers as possible.[12]

The second force—in a sense a counterforce—is the competing facilities. While enterprises seek to find in the existing trading-area network a "gap" large enough for successful investment, such "gaps" cannot always be found. Therefore, a forward-looking enterprise may endeavor to create a new arrangement of networks in the market area. This is done by seeking new locations on modern transportation arteries, more frequently focal, but at times eccentric, to the existing population. The site requirements generally can be found only on highway locations and large tracts of land suitable for the new type of store, with its superior facilities and merchandising techniques. This new type of store generates enough merchandising power to create a new trading area of considerable magnitude. As the number of such new store trading areas is multiplied, a new network of trading areas is created.

As stated, the problem of finding "gaps" in the existing arrangement is complex. One company may consider a gap to exist when 20 percent of an entire market area is not served adequately by existing store facilities. Another company, however, with a policy of building high-volume stores, may not consider that a gap exists in this market, because the available "surplus" (potential) is insufficient to meet this company's store-volume needs.

What compounds the problem is that a company with a number of established stores in a given market is certain to lose ground in some of its locations, as competitive conditions change. Hence, it must be growing elsewhere or it faces decline, and even possible disaster. A business enterprise must be able to sustain dynamic growth from its own resources, by generating profits. Without dynamic growth, there can be no survival. Dynamic growth means keeping pace with market opportunity changes; it also means changing physical facilities to keep pace with competition. A business, therefore, must rejuvenate itself by replacing its outmoded facilities and must increase in size through building additional facilities.

TABLE 4.—GENERALIZED EXAMPLE OF TRADING AREA COMPONENTS BY PER CAPITA SALES

Store type	Annual sales volume	Annual per capita sales boundaries		
		Primary trading area	Secondary trading area	Fringe trading area
Department store	$12,000,000	$30	$15	$3
Discount store	8,000,000	20	8	2
Supermarket (in large shopping center)	2,500,000	50	15	4
Junior department store	2,000,000	13	5	2

The growth of a business enterprise can take place by expansion into completely new territories, and by filling existing gaps or rearranging trading-area networks. There is a distinction, however, between, on the one hand, creating new store trading areas that mesh in with existing trading areas, and, on the other hand, creating new store trading areas that overlap or lie completely within existing store trading areas. Under the former condition, store trading-area networks are filled in "harmoniously"; under the latter, complete "surgical" rearrangements occur.

Where, for example, a store is built at the juncture of two separate store trading areas, and this store is comparable to its surrounding competitors, the new store adds a third, intervening, trading area which overlaps segments of the other two trading areas and becomes part of the existing network. Figure 12 illustrates examples of both complementary store trading-area networks and new ones that rearrange the network. This map shows how a city of 100,000 population has been divided into a number of discrete store trading areas by 30 supermarkets. The most recently built supermarkets have not moved into gaps within existing trading areas; nor have they complemented the trading-area network by taking

[12] August Lösch writes, "The market networks for nonagricultural enterprises are widely spread out in space . . . 'belts' disturb our theoretical economic geometry much less than districts do, for example." Lösch, *op. cit.*, p. 372.

Edgar M. Hoover states: "There is an internal contradiction in Lösch's theory. The hexagonal market (or supply) areas that he made assume a uniform density of demand (or supply); but once he has established his concentrations of production, demand and supply are thenceforth not uniformly distributed any more and the hexagons ought to be of different sizes and shapes." Correspondence, December, 1959.

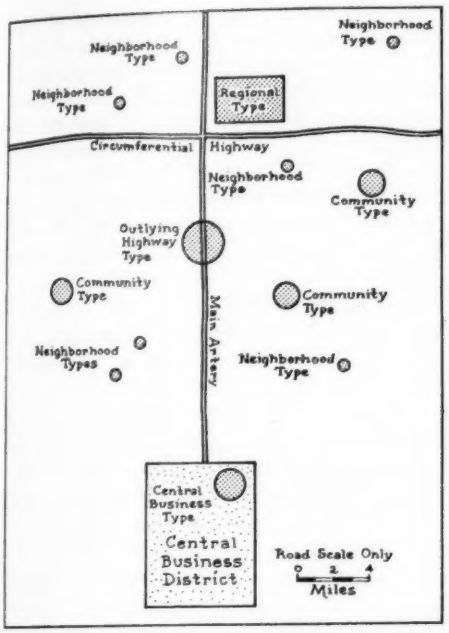


FIG. 13. Schematic presentation of the location of planned shopping centers in metropolitan areas. In examining existing planned shopping center patterns, it was found that the outward side of regional shopping centers generally supports only small, not medium-size centers.

up intervening locations. Instead, they have quadrupled their size, have introduced in some instances a complete line of nonfoods, and are in the process of rearranging the trading area network.

Another type of rearrangement occurs in smaller cities and towns when important new retail facilities are developed at the periphery of the town. Thus, whereas formerly the town had only one trading area tributary to the CBD, the new arrangement results in additional and overlapping trading areas.

New sets of small trading areas are created within existing larger trading areas by establishing stores with modern facilities to capitalize on special customer demand for convenience of location (see Fig. 12, stage 3).

The hierarchy of store trading-area networks in the Anglo–American city has been made more complex by the spectacular post-World War II growth of suburbia and the advent of the planned shopping center. Utilizing the classification presented earlier, Figure 13 presents this hierarchy schematically. This diagram indicates that while community and neighborhood planned shopping centers are found between regional shopping centers and the CBD, usually only neighborhood planned shopping centers are found on the outward-facing side of the regional shopping center. This development results from the sparser population density on the outward-side of the regional shopping center and, correspondingly, also fewer "interceptor" locations.

Two remarkable facts about planned shopping centers are: (1) the ease and speed of their emergence within the market area, and (2) their impact upon new and existing shopping facilities. Either real estate developers or retailer-owned real estate enterprises are the prime movers responsible for the planned shopping center. These are men and firms with force and determination who have promoted and effected concentrations of modern shopping facilities almost overnight.

These new shopping-facility concentrations may have, depending perhaps on one's point of view, either good or bad effects upon the retail structure of a city, but certainly the effects are profound. The impact of new planned shopping centers on older store facilities, particularly the CBD, varies considerably with three factors: (1) growth of population and the corresponding need for new store facilities, (2) relative amount and kind of new store facilities provided by planned shopping centers, and (3) response and countermeasures taken by merchants operating existing older store facilities to meet the new competition. The relative importance of the CBD's share of the total retail sales in metropolitan areas is adversely affected by the competition of the planned shopping centers.

Many of the planned shopping centers house stores with higher- and medium-priced goods, which appeal mainly to the middle and higher income groups of their trading areas. In these instances, lower-priced stores or strips of stores tend to locate near these planned shopping centers to tap essentially the lower income group of the same trading area.

The planned shopping center can upset equilibrium as it competes with older facilities. But it can also upset equilibrium by forcing retailers to concentrate new facilities rather than to spread them uniformly. In smaller cities, in particular, leading retailers may not

be enthusiastic about a planned shopping center, but they also may not dare to stay out of the development. Furthermore, if a developer insists on having two stores of the same type in a shopping center, then even though an operator may prefer another location or may desire to wait for further population growth, he may nonetheless enter the shopping center to keep a competitor out. Thus, overbuilding of stores or poor distribution of retail facilities within the market may result.

In general, the larger the retail district, the greater the percent of retail sales to total sales obtained by shopping-goods establishments in that district; the smaller the retail district, the greater the percent of sales obtained by convenience-goods stores. The pattern of store trading-area networks is one in which the CBD trading area has been superimposed upon the entire metropolitan area. Secondary business districts attempt to compete with the CBD, and their smaller trading areas lie within that of the CBD. The string developments that adjoin the CBD, with their heavy emphasis on automotive and furniture stores, duplicate the CBD in the total coverage of the metropolitan area, just as strips adjoining the secondary business districts duplicate the latter.

Within this framework are the convenience neighborhood or isolated stores whose small trading areas form islands within the trading areas of greater magnitude.

While planned shopping centers of various sizes generally exhibit the same hierarchical patterns of trading areas as do unplanned business districts, there are important deviations. Essentially these deviations result from the types of stores found in the planned shopping centers.

Planned shopping centers offer advantages of proximity to home, ease of parking, minimum traffic friction, and space for large stores. Here the customer can shop comfortably not only for food and other convenience goods but also for a wide selection of general merchandise. Thus, food stores may account for only 5 percent of the total sales within a large CBD, but for as much as 20 percent of the total sales of a regional planned shopping center. Shopping-goods stores, on the other hand, may account for only 10 percent of the total sales of a neighborhood business district and 20 percent of total sales of a neighborhood shopping center.

The trend to a higher percentage of convenience-goods sales in large planned shopping centers and to a higher percentage of shopping-goods sales in small planned shopping centers can result in a more uniform distribution of planned shopping centers than now exists. Furthermore, the altered association of stores in the planned shopping centers may effect a greater coincidence of individual store trading areas for the different stores within a given shopping center.

Few planned shopping centers have been built in association with other business activities—large numbers of offices, wholesaling and manufacturing plants. Consequently, even the largest of the regional planned shopping centers cannot exert a power of penetration commensurate with that exerted by the competing CBD, in superimposing its trading area upon the trading areas of other planned shopping centers of lesser magnitude. Thus, while the trading area of a large planned shopping center is superimposed upon that of a smaller planned shopping center, there is not a complete overlap for all stores. Figure 14 shows this. Here a full-line department store exists in the former but not in the latter shopping center. For the supermarket of the larger shopping center, however, the existence within the smaller shopping center of another modern supermarket belonging to the same company limits its power to superimpose its trading area.

STORE SATURATION AND MARKET EQUILIBRIUM

Store Saturation

Store saturation of a market is a condition under which existing store facilities are utilized efficiently and meet customer needs adequately. When a market has just enough store facilities of a given type to serve the population of the market satisfactorily and yield a fair return to the owners on their investments, without raising prices to the customer to achieve this return, then saturation exists for the given type of store. When a market has too few stores to provide satisfactorily the needs of the customer, it is *understored.* When a market area has too many stores to yield a fair return on investment, it is *overstored.*

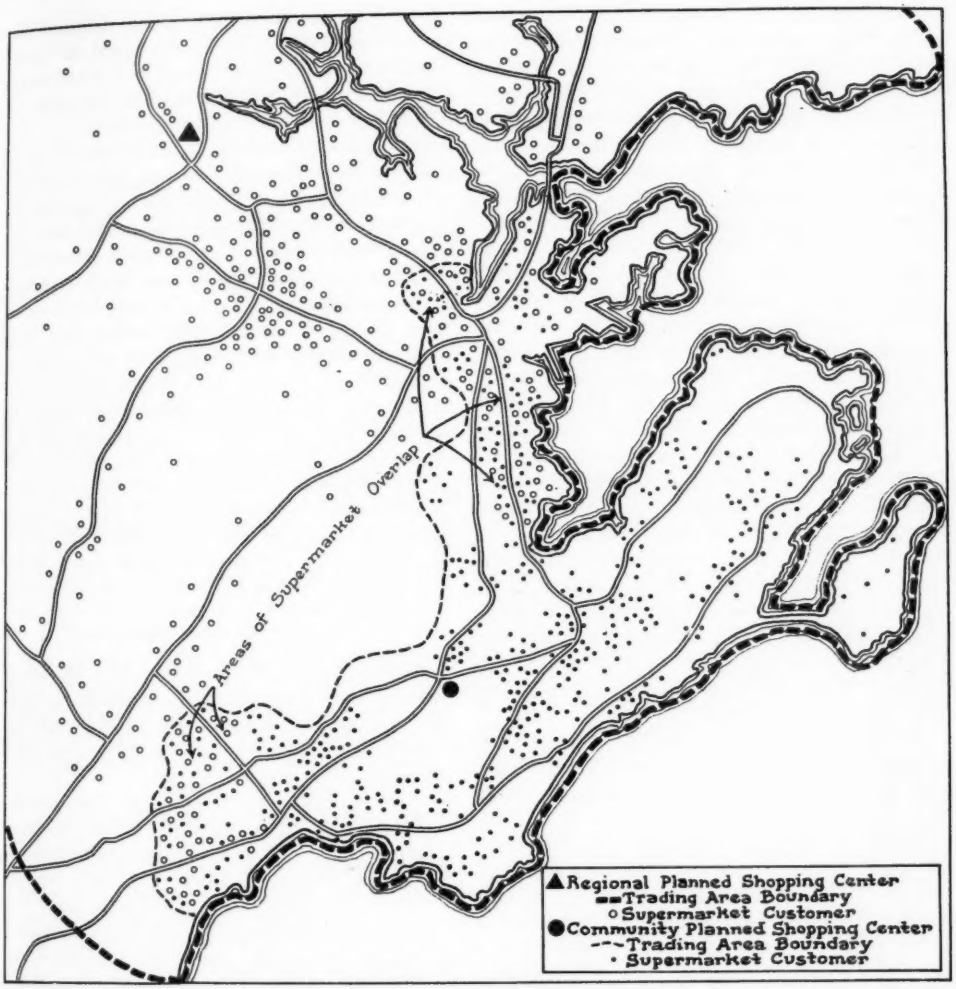


FIG. 14. Variation in overlap of store trading areas. While there is trading area overlap for planned shopping centers of differing sizes and functions, this overlap does not apply to all stores within the centers. Two modern sister supermarkets, one located in a regional planned shopping center and the other in a smaller center show a very limited overlap in their respective trading areas.

Static and Dynamic Equilibrium

Store saturation is an ephemeral condition, for balance between facilities and their use cannot be rigidly maintained in an inherently dynamic situation. Additional facilities or changes in merchandising techniques or in the geographical environment will upset the balance.

The tendency is for stores of the same type to equal each other in size insofar as market opportunities permit. Such equality in size becomes a competitive offensive and defensive necessity to satisfy customers and to achieve optimum operating and/or merchandising efficiencies in a given marketing environment. If all companies expand store facilities beyond the needs of the market, then a situation is created where all have lowered their productivity in terms of sales per square foot of selling area.

Disturbance in equilibrium may occur if anticipated population increases fail to material-

ize or if an area is depopulated. It may take place if there is a basic change in the transportation pattern, opening the existing trading area to outside competition. It may occur if customer shopping habits, affected by mode and frequency of travel, alter radically.

If the balance between need and availability of store facilities is restored without appreciably altering the existing trading-area pattern, then the period of imbalance can be described as one of "short-run disequilibrium." Such a restoration of balance occurs in situations when understoring is corrected through the building of additional store facilities to keep pace with population growth. This type of restoration may take place in fast-growing suburban areas where subdivisions of a few thousand homes may spring up overnight. It may also take place in mature areas, such as the inner portions of central cities which become engulfed by the immigration of younger persons with large families, for example, in cities in the northern United States which receive annually thousands of immigrants from the South. Such restoration can occur under conditions of overstoring through growth of population, or through the elimination of some unprofitable facilities.

At times, however, the balance between the availability of and need for store facilities can only be restored through innovations in facilities, through vastly improved uses of existing facilities, or through changes in merchandising techniques. These revolutionary changes usually seek lower operating costs, and/or greater customer satisfactions—in the form of lower prices, a more pleasing shopping environment, or increased services. The period of imbalance that marks such changes is one of "long-run disequilibrium."

As innovations in facilities and merchandising techniques are made (see Fig. 12), networks of new trading-area patterns are introduced. This process was set in motion as: (1) grocery stores gave way to supermarkets, movie houses to drive-in theaters, corner gasoline stations to super-service stations, dry goods stores to self-service general merchandise stores, and drug stores to super-drug stores, and (2) planned shopping centers emerged in competition with older business districts.

When saturation or balance is restored under short-run conditions, the equilibrium that is established might be described as "static equilibrium." When balance is restored under long-run conditions, the equilibrium might be described as "dynamic equilibrium."

In viewing the pattern of the location of retail facilities through a system of equilibrium, we consider the following among the disturbing factors that constantly operate within the environment to upset equilibrium: changes in population, accessibility, merchandise assortment, purchasing power, demand, retail facilities, shopping and buying habits, and concepts of profitability margin or return on investment. These changes bring about conditions representing either a lack or an excess of needed store facilities. The periods of imbalance, or disequilibrium, are lengthy and saturation, the point of equilibrium, is brief in time and in space. Yet it is equilibrium which is the basis for an orderly arrangement of patterns of retailing facilities, for it is the constant goal (see Fig. 15).

How to Measure Saturation

Measurement of saturation is a matter of professional interest to economists, geographers, and city planners, and of practical concern to businessmen. Several approaches may be taken:

1. Number of persons per retail establishment. The national or regional norm of the number of persons per specific type of store is calculated from available census or other statistical data. Similarly, the number of people for the same type of store is calculable for the given market; the figure thus obtained is compared with the national or regional norm to measure saturation.

2. Available retail expenditures per retail establishment. The number of persons times per capita retail expenditures for specific types of goods is calculated for the given market. The figure obtained is compared with national or regional averages. There are many difficulties in using this measurement: No allowance is made for differences in store size. Average purchasing power disguises too many relevant factors. Different income groups may buy different things or in different amounts. People buy different goods at different times in their lives; thus the same number of toy stores serving a

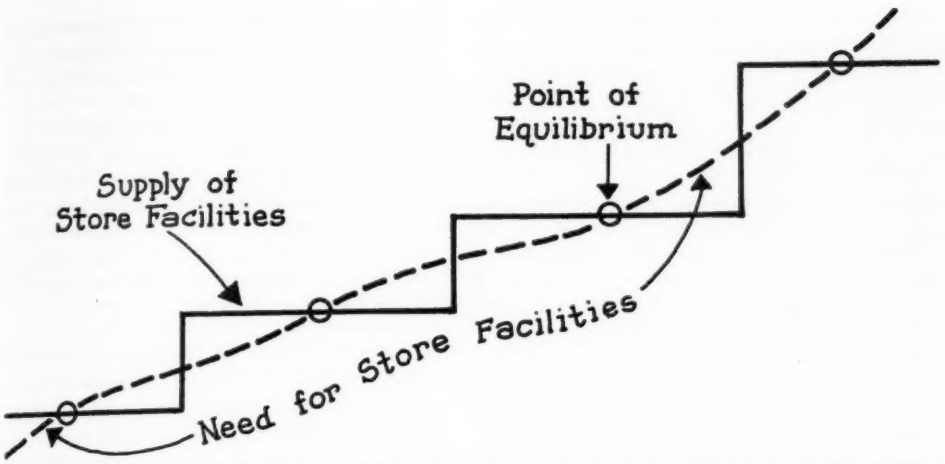


FIG. 15. Schematic presentation of store equilibrium. Equilibrium is the basis for an orderly arrangement of patterns of retailing facilities, for it is the constant goal. Here equilibrium is presented schematically as the point of balance between the need for store facilities and the supply of store facilities.

youthful population will provide an excess of facilities to this population as it ages.

3. Number of persons per store-front footage. The data for a given market are compared with averages determined in other studies, made mostly by city planners. This does not take into account multiple-story facilities. Store frontage without store depth is not a true index of store area, and city blocks differ in depth. Frequently, in older areas stores expand their frontage but do not expand their depth correspondingly; whereas in newer areas buildings can be extended in depth more readily.

4. Ratio of store vacancies to total stores. This ratio can be measured on the basis of number of vacant store or vacant floor area. Many vacancies imply overstoring. However, lack of vacancies may not mean the converse, because many stores do not use their space fully or effectively.

5. Sales per square foot of selling area. This, in our opinion, does measure productivity and efficiency of use of space. Therefore it is suggested as the basic measure of saturation. (The use of sales per square foot of total store area gives lesser comparability than sales per square foot of selling area.)

There are capacity norms of selling-area productivity for various types of retail business. Related to these sales-capacity norms are rentals and other operating costs. For example, supermarket operators are not likely to establish a store at a given location if estimated annual potential sales per square foot of selling area are not likely to exceed $150 or even more. Many department-store operators are not likely to build new facilities if these facilities cannot generate estimated annual sales of at least $100 per square foot of selling area.

As with all standards, there are difficulties in using the measure of sales per square foot of selling area. Some of these difficulties are:

1. Differences in company policies, among enterprises within the same type of retailing, result in variations in the percentage of total store area devoted to selling area. For various operating reasons, some companies devote more store space than others to storage, preparation of goods, office use, etc.

2. Changes in price index require statistical adjustments of per capita sales figures for comparability.

3. Differences and changes in the merchandise mix (i.e., variety and quality of type of goods) handled also require weighting of measurement standards. If one store devotes 15 percent of its space to slow-moving merchandise, and a second store only 5 percent, then sales per square foot of selling area are not strictly comparable

unless adjustment is made for the difference. Quality differences present a similar problem. In any case available information may not be adequate to avoid some margin of error.

4. Because of differences in number of hours per week that stores are open for business, some stores use their selling facilities more intensively than other stores. While sales productivity per square foot is stepped up by such more intensive use of store facilities, normal sales become more spread out or diffused and the additional operating expenses incurred may exceed the gains from higher sales per square foot. To make meaningful measurement adjustments for these differences is a complicated task.

5. Finally, there are differences in merchandising methods. Self-service merchandising generally means a more rapid turnover of goods and higher sales per square foot of selling area than does service merchandising. Mail order business can be done from a desk with little or no merchandise on the premises. Similarly, sales to customers over the telephone require little selling area. Judicious allowances must be made for these complicating factors in using sales per square foot of selling area as a measurement of store saturation.

In spite of these difficulties, sales per square foot of selling area is the most significant broad guide to store productivity that we can find. The smaller the sales area in relation to sales:

(1) The less the investment in land, building, and fixtures.
(2) The less the inventory on display and hence the greater the turnover.
(3) The fewer the steps for employees.
(4) The less the time that is consumed by customers in completing their purchases.

Thus, higher sales per square foot of sales area means greater return on investment. If sales per square foot of selling area are unusually high, and where customers are inconvenienced in travel to shop (too far) and while shopping (too crowded and slow), then a situation of understoring exists. Where the reverse is the case, overstoring exists. Thus, while supermarkets with annual sales of $150 per square foot of selling area represent the norm for new stores, sales of $300 per square foot would indicate understoring and sales of $100 per square foot, overstoring.

Attempts have been made to measure productivity on the basis of profits per square foot of selling area, or on the basis of potential sales per square foot of selling area. The former, however, requires too many adjustments to make the measure meaningful for comparative analysis; the latter is unrealistic for practical application.

STORE SATURATION AT VARIOUS ORGANIZATION LEVELS AND FOR VARIOUS MARKET AREAS

Understoring, saturation, and overstoring for a given type of store occur on two levels: for the group as a whole, and for specific companies within the group. Thus, in a given self-contained market area of 500,000 persons, there might be 400,000 square feet of department store selling area, and annual sales might average $125 per square foot. Since $100 per square foot of selling area is the generally accepted capacity norm for department stores, the market area could be regarded as understored.

If, however, within this market area one company already accounts for one-quarter of the total department store sales, then the company might consider the market saturated with respect to itself. Few department store companies seek more than 30 percent of the total market in metropolitan areas of this size.

For each company, there is a diminishing return on investment if it seeks to obtain too great a share of the market. When this occurs, competition reacts through expansion of facilities or stronger merchandising. The company in question, then, to hold its share of the market, must increase services or lower prices, thus reducing its return on investment.

Saturation, as expressed by share of market goals for individual companies, varies with the strength and philosophy of the company and with the general market area served. In larger metropolitan areas, some chain department-store companies set 20 percent of the market as saturation, while in smaller metropolitan areas saturation is set at 30 percent. Similarly, chain supermarket companies set their saturation goals for nonmetropolitan areas at 25 to 30 percent of the market, and

at 10 to 15 percent of the market in large urbanized areas. This difference springs from two considerations: (1) The amount of sales needed to support a modern store facility in nonmetropolitan areas can be achieved only by getting large per capita sales. (2) In metropolitan areas there are usually a number of strong companies competing with each other and the sales are divided more evenly among them.

Within a given market, saturation for individual companies can vary with company policy. In some cases, saturation is established by set minimum goals of store-sales for new stores. If the gaps between or within existing trading areas do not promise such minimum sales, then the market is overstored for this company. One national chain in the United States does not establish a new supermarket unless estimated sales are at least $1,000,000. A leading regional chain has set $2,000,000 as its minimum of new supermarket annual sales. An aggressive local chain seeks minimum sales of $3,000,000 for its new supermarkets. Thus, areas that are overstored for the last company may not be overstored for the first two. Furthermore, if the local chain decides to change its policy on store sales goals, the market may offer it opportunities for opening smaller supermarkets. On the other hand, the store "personality" of the aggressive local chain may be powerful enough to attract $3,000,000 annual sales, in part from existing stores operated by the national chain and by the regional chain, in spite of the overstoring.

A supermarket company may be willing, because of a desire to add store facilities, to lower its yardsticks for saturation in another way: it may set saturation at $125 per square foot of selling area, rather than the industry's capacity norm of $150. This desire may be based upon a number of considerations:

(1) To gain better purchasing terms.
(2) To provide or make better use of warehousing facilities.
(3) To provide more retail outlets for its own processed goods.
(4) To reduce relative costs of advertising or store supervision.
(5) To anticipate future store space requirements for handling additional lines of goods.

Scrambling of Merchandise

If all retail stores sold the same goods in the same manner, the conditions for market equilibrium would be met by small, compact trading areas surrounding each store. The stores would be uniformly spread, insofar as the distribution of population and lines of transportation would permit. Of course, instead of a more or less uniform spread of stores, all of these general retail facilities could be clustered in a compact belt. But this has not happened. Rather than serving as general stores for all goods, retail stores have specialized along particular lines, benefiting from specialized facilities, merchandising skills, and operational techniques.

The customer's choice of specialized offerings of goods versus generalized offerings is the basis for store trading area differentiation. The fewer the number of stores that carry the same item, the more extensive are their trading areas. The less frequent the number of visits needed to replace an item, the more extensive is the trading area from which these customers come. Conversely, the greater the number of stores that carry the same item, and the more frequent customer visits are needed to replace these items, the less extensive is the trading area.

For a number of years now there has been a tendency in the United States for the specialized retail outlets to scramble their merchandise mix by taking on lines traditionally sold by other types of store. Thus self-service dry goods stores have taken on appliances; super-drugstores have extended lines to include variety, hardware, and general merchandise; supermarkets have added health and beauty aids, housewares, and garden supplies, and are now expanding into soft goods.

In some instances, stores do extend their trading areas by scrambling their goods. On the other hand, by scrambling goods, stores may concentrate more of their sales in nearby areas and decrease their drawing power to the more distant parts of the trading area.

Two forces intensify the scrambling of merchandise: (1) economy of scale derived from larger buildings, larger parking lots, and a wider variety of goods that can be sold under one roof, and (2) the planned shopping center, which stimulates other stores to try to duplicate all of the one-stop shopping offered

by the shopping center. The granting of "monopoly" positions to national and regional chains by American shopping-center developers, at the expense of independent merchants or small chains, has speeded the development of large combination food and general merchandise stores and all-purpose discount stores adjacent to shopping centers or on the fringes of their trading areas. This is quite different from parts of Europe where independent merchants are so firmly entrenched as to restrain the development of large chains.

Scrambling of merchandise occurs also in stores located within planned shopping centers through duplication of lines carried. There are many instances where half a dozen stores within a shopping center will devote space to the same types of goods, frequently of the same brands or varying little, if any, in quality and price. The element of association with successful neighbors, and the desire to secure a larger share of the customer's expendable income are reasons for this.

There are, of course, such limitations to scrambling as cost of operation, space, merchandise knowledge, and restriction or other means of regulation (liquor permits, zoning ordinances, lease restrictions to the sale of goods).

Scrambling of merchandise also leads to wider selection, more attractive pricing and better presentation than does specializing. But in the long run, if scrambling is to lead stores outside shopping centers to duplicate shopping centers, or shopping centers to try to duplicate one another by appealing to the same types of customers through the same lines of medium- to high-priced merchandise, then store locations, as a whole, will take on a more uniform distribution pattern.

Competition and Market Equilibrium

A competitor who miscalculates and overbuilds can create disequilibrium. A powerful competitor who overbuilds can ruin others. A weak competitor who miscalculates can ruin himself. However, it is not necessary to have new store facilities built to touch off conditions leading to disequilibrium. A change in store ownership, and with it in store personality, can stimulate others to add new facilities. The absorption of a strong local chain by a larger company can result in a weakening of the personality and performance of the absorbed stores, giving competitors an incentive to expand.

Mergers are frequently considered a vehicle for resolving or averting problems of overstoring in a market area. Merging works well if the store facilities added by the acquiring firm can be used efficiently or partly disposed of; otherwise, a merger, rather than restoring equilibrium, can heighten a condition of disequilibrium.

A change in ownership may disrupt equilibrium, not because the new owner is weak, but because some of the added facilities cannot be used effectively. Because competitive stores are being located within shopping centers, and close to each other on highways, or because bigger stores are being built with more extensive trading areas (and more overlap with competition), it becomes increasingly difficult for one chain to absorb a second without acquiring duplicatory or otherwise inefficient store facilities.

New building programs and/or reshuffling of store ownership result in a change in the balance between available store facilities and their use. Building programs can be either boons or detriments, and there is never a perfect merger.

Conclusion

Changes in store facilities, merchandising, and services are the means by which dynamic equilibrium is maintained by a retailing society. Addition to existing facilities is the method by which static equilibrium is maintained.

In general, the more regulated the retailing society, the greater the possibilities of freezing store facilities and trading areas. The more competitive the society, the more likely it is for change to be introduced, shaking up the equilibrium and hastening the process of store adjustments.

Government and political influences can be put to positive use. They can serve to induce or anticipate change. In the United States, zoning and licensing may prevent builders from bringing about changes, but they can also serve as vehicles for encouraging the development of additional shopping facilities where needed. In the English "new towns," the planning process is used to anticipate changing needs.

Acting as regulators or breaks are capital limitations and the high investment cost in modern retail facilities. On the other hand, the basic value of land, particularly in inflationary periods, as compared to the more ephemeral value of buildings, makes it easier to shift retail store facilities to new sites and use older sites for other purposes.

To prevent run-away changes in store facilities within a competitive society, sound planning by individual companies is essential. Such planning must undertake to evaluate the potential needs of the market area as well as anticipate store location moves by competitors. Under rational behavior of retail society, strong and well-informed competition is the best agent for maintaining dynamic equilibrium within market areas.

THE MARINE TERRACES OF ARUBA, BONAIRE, AND CURAÇAO, NETHERLANDS ANTILLES

CHARLES S. ALEXANDER
University of Illinois

ARUBA, Bonaire, and Curaçao, the leeward islands of the Netherlands Antilles, are rimmed by several clearly defined marine terraces which, throughout much of the length of the islands' coasts, are cut into coral limestone (Fig. 1).[1] The prominent shape of the wave-cut terraces is in part the result of the position of the islands in the arid belt that extends along the southern shore of the Caribbean. The resistance of limestone to arid-climate weathering has kept the terraces in an excellent state of preservation, while their surfaces are only slightly obscured by the light covering of xerophytic vegetation (Fig. 2).

PREVIOUS TERRACE STUDIES

These well-defined terraces have received the attention of several geologists. Martin,[2] in his work on the Netherlands Antilles, was the first to describe the terraces on Aruba, Bonaire, and Curaçao. He recognized one terrace on Aruba, three on Bonaire, and four on Curaçao, and interpreted all of them as uplifted platforms of marine abrasion.

In his discussion of the geology of Aruba, Westermann described two Pleistocene limestone deposits.[3] He identified the older and higher of these as the remains of a limestone cap that formerly covered most of Aruba, while the younger and lower occurs as limestone terraces. This lower deposit is of varying width, and with minor interruptions extends around the island. On the southeast coast of Aruba this terrace limestone is separated from the higher, older limestone by an escarpment. Along the southwest coast, however, the two deposits are not separated by an escarpment but merge gradually inland. The younger limestone is not everywhere limited to a single terrace. Near Rincon on the southeast coast it occurs in two low terraces separated by an escarpment (Fig. 4). According to Westermann, the younger limestone may have been deposited during the rise in sea level following the (last?) glacial period. An upheaval of the land followed this postglacial transgression, exposing the younger Pleistocene limestone. A second small uplift then resulted in the formation of a second terrace in some localities. A "sub-recent" lowering of sea level (c. 15–20 feet), Westermann reasons, exposed a few very young limestone deposits (too small to be shown at the scale of Fig. 4). This postulated drop in sea level also brought to the surface the coral barrier reef along the southwest coast of the island (Fig. 4).

Pijpers, in his work on the geology of Bonaire, substantiated Martin's report of the number of terraces.[4] He cited the varying height of the terraces and the fact that some of them merge horizontally as indications of unequal uplift of the island during the Quaternary. Recently Westermann and Zonneveld postulated the presence of seven marine terraces on the north and east coasts of Bonaire.[5] However, as the result of detailed field work on the entire north and east coasts, it is my considered judgment that several of the levels mentioned by Westermann and Zonneveld

[1] The field work on the islands, supported by a grant from the Office of Naval Research, was carried out during the summers of 1955 and 1956. Additional funds were made available by the University of Illinois Research Board to cover certain research costs both in the field and at home. Black-and-white reprints of original Dutch topographic maps, made by the Army Map Service, were used in the field work. The map scale is 1:20,000 for Curaçao and Bonaire and 1:40,000 for Aruba. The contour interval on all these maps is ten meters.

The maps and diagrams used in this paper are the work of Don R. Hoy, research assistant, and James A. Bier, cartographer, both of the Department of Geography, University of Illinois.

[2] K. Martin, *Bericht über eine Reise nach niederländisch West-Indien and darauf gegründete Studien*, Part II, *Geologie* (Leiden, 1888), pp. 106–33.

[3] Jan Hugo Westermann, *The Geology of Aruba* (Utrecht, 1932), pp. 123–4.

[4] Petrus Johannus Pijpers, *Geology and Paleontology of Bonaire (D. W. I.)*, (Utrecht, 1933), p. 44.

[5] J. H. Westermann and H. I. S. Zonneveld, *Photogeological observations and land capability and land use survey of the island of Bonaire (Netherlands Antilles)*, issued by Koninklijk Instituut voor de Tropen (Royal Tropical Institute), Amsterdam. Mededeling No. CXXIII. Afdeling Tropische Producten No. 47 (1956), p. 21.

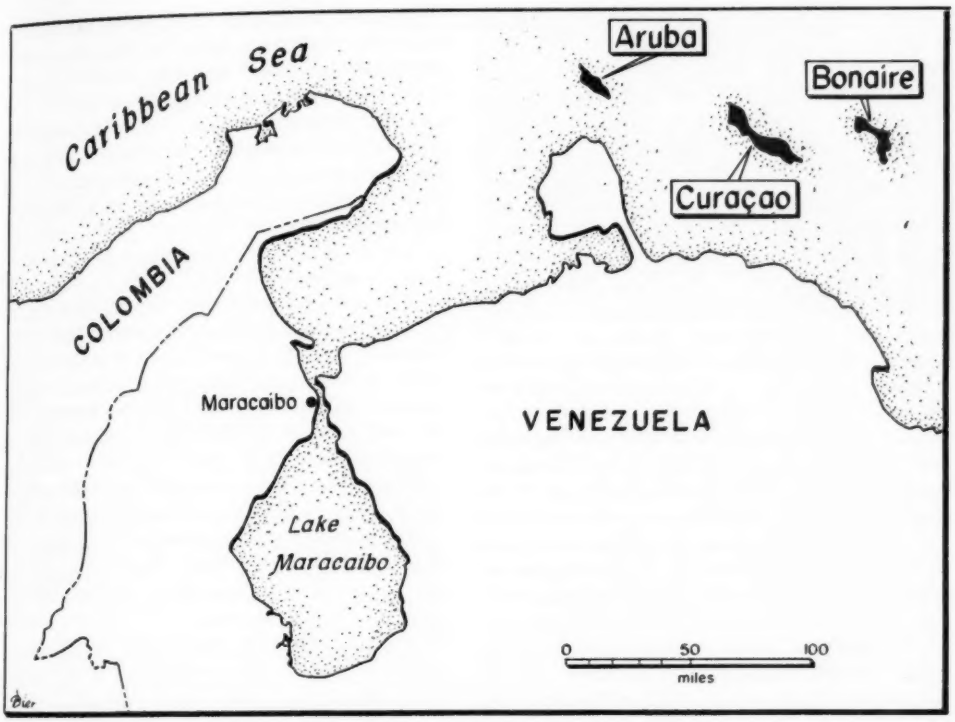


FIG. 1. Leeward islands of the Netherlands Antilles.

are not marine terraces. For example, the one-meter level could be better explained as a solution bench. The one they postulate at 5–6 meters may be the terrace-like feature at that elevation in the vicinity of Boca Onima. Several small normal faults occur here at this level, and since the down-dropped block lies to seaward of the fault scarp, they obviously account for the localized terrace-like form.

Molengraaf believed that most of the terraces of Curaçao were the result of intermittent uplift of the island.[6] However, he designated one terrace as resulting from a eustatic rise in sea level. He believed that following the last Pleistocene glaciation the sea level rose and cut a terrace in the limestone at the present 230-foot contour, destroying parts of the uplift terraces. Then, he speculated, Curaçao was uplifted about 210 feet to nearly its present position. A pause in this uplift resulted in a marine terrace (the first terrace), which now has an elevation of from 50 to 65 feet. Finally, to explain the horizontal beds of coral and shell around the sides of the bays, he suggested a sinking of sea level by about 20 feet.[7]

FIG. 2. The third marine terrace near Hato on Curaçao. The backslope in the background is about 50 feet high.

[6] G. J. H. Molengraaf, "De Geologie van Nederlandsche West-Indie: Curaçao," *Leidsche (Leidse) Geologische Mededeelingen* Vol. 5 (1931?), pp. 684–85.

[7] *Ibid.*, p. 687.

Recently P. H. Buisonjé of the Geologisch Instituut, Amsterdam, investigated the Quaternary deposits of Aruba, Bonaire, and Curaçao. The results of this study have not been published.

GEOLOGY

The three islands are quite similar from a geologic standpoint.[8] The basement of each island consists of volcanic diabase and sedimentary rock of Cretaceous age, which were folded and, especially in Aruba, injected by diorite magma during the early Tertiary. Overlying the basement rocks are limited occurrences of Tertiary limestone and rather extensive deposits of Pleistocene limestone. On each island there are unconformities between the Cretaceous and Tertiary deposits, and between the Tertiary and Pleistocene rocks. All three islands, then, appear to be areas of crustal instability, each showing evidence of several periods of uplift. The most recent uplift began in the early Pleistocene following a period of relatively higher sea level during which a cap composed mainly of coral limestone, but (especially on Bonaire) including aeolian deposits of calcareous sand, was deposited over much of the older rocks of the islands' formations. This last uplift, according to Westermann, was accompanied by an arching of the cap and by subaerial erosion of the limestone and underlying rocks.[9] The uplift was also accompanied by the erosion of several marine terraces into the flanks of the islands and the deposition of younger coral limestone on the terrace platforms.

Schaub is of the opinion that the arching of the limestone on Curaçao is the result of initial dip; and that the cap was not deformed as the island emerged from the sea.[10] Schaub does not distinguish whether the emergence represented a movement of the land or of sea level.

Of the three islands only Aruba has a barrier reef. The reef generally lies about 400 yards offshore and extends discontinuously along the southwest coast between Punt Basora and Oranjestad (Fig. 4). It is composed of coral and is capped by a beach ridge of coral fragments 2 to 10 feet in height. On the seaward side, the reef extends to a depth of at least 150 feet.

MARINE TERRACES

Originally marine terraces were thought to develop under conditions of intermittently rising land and stable sea level. A more recent view considers the land to have been stable and the sea falling and rising in response to the waxing and waning of Pleistocene continental glaciations.[11] In the latter case the superposition of the terraces is believed to result from a progressive lowering of sea level, which occurred throughout the Pleistocene. A combination of both rising land and changing sea level has received little attention in the study of marine terraces.

In the earlier studies of the terraces on the three islands, only the condition of an intermittently rising land with a stable sea level has been considered, although the circumstance of a fluctuating sea was sometimes added in the discussion of the origin of the youngest terrace. In this paper I wish to investigate the possibility of rising land and changing sea level occurring together to form marine terraces.

In the course of the field investigation necessary to carry out this investigation the terraces of the three islands were mapped and their elevations were carefully measured at points as close as possible to the back edge.[12] Elevation traverses across the terraces were made with a hand level from points of known elevation or, where these were lacking, from the water's edge. (The tidal range around

[8] L. M. R. Rutten, "De Geologische Geschiedenis der drie Nederlandsche Benedenwindsche Eilanden," *De West-Indische Gids*, Vol. 13 (1931–32), pp. 401–41; J. H. Westermann, *Overzicht van geologische en Mijnbouwkundige Kennis den Nederlandse Antillen*, issued by Koninklijke Vereeninging Indisch Instituut. Mededeling No. LXXXV (1949), pp. 133–34; Westermann and Zonneveld, *op. cit.*, p. 21.

[9] Westermann, *Overzicht . . . , op. cit.*, p. 134.

[10] H. P. Schaub, "Geological Observations on Curaçao, N. W. I.," *Bulletin*, American Association of Petroleum Geologists, Vol. 32, part 2 (1948), pp. 1, 276–83.

[11] Frederick E. Zeuner, *The Pleistocene Period* (London, 1959), pp. 276–307.

[12] The back edge is the line on the landward side of the terrace defined by a wave-cut notch at the foot of the terrace blackslope or old sea cliff, or by the intersection of the backslope and the abrasion surface. The abrasion surface is the wave-cut surface of the terrace. The terrace platform differs in that it includes both wave-cut surface and the overlying mantle of marine and subaerial deposits.

the island is one foot or less).[13] To reduce error on long traverses, intermediate measurement stations were set up. Approximately midway in the traverse a prominent landmark was selected, or, lacking that, a marker was erected. Several measurements of the elevation of the station were then made, and the average value of these was assigned as its elevation. This elevation was then also used as a base for making measurements at higher elevations. At least two, and sometimes three, measurements were made of the elevation of each station and of the back edge of the terrace. The values of the elevation of the back edge of each terrace were averaged, and the resulting means were used as the terrace heights. Occasionally an elevation measurement at a particular place would be appreciably different from others made there. This value was discarded and the measurement was redone. Frequent checks on the accuracy of the hand-level survey method were made against points of known elevation. These measurements, when averaged, indicated the possibility of maximum error in the terrace elevations ranging from plus or minus 5 feet to plus or minus 10 feet, the greater error occurring on longer traverses. Usually the error was less, plus or minus 2–3 feet in short traverses, 3–6 feet on longer ones.

Other errors in determination of height may be caused by alluviation at the foot of the backslope and by solution weathering, which in some cases may lower the terrace platform from 4 to 6 feet.[14] Areas of alluvial deposition could be easily recognized and avoided, and it soon became evident that pronounced solution weathering was confined to localized areas of concentrated runoff. Consequently, errors in elevation due to subaerial erosion and deposition could be minimized by careful selection of measurement sites. Widespread occurrence of a wave-cut notch at the base of the terrace backslope aided in the selection of appropriate points for the measurement of elevation. These wave-cut notches occur most frequently on the northeast sides of the islands and are especially well preserved on the lower terraces (Fig. 3).

FIG. 3. A wave-cut notch at the foot of the backslope of the first terrace near Hato on the windward coast of Curaçao.

Field investigation showed that the marine terraces of Aruba, Bonaire, and Curaçao are best developed on the windward, or northeast, coasts. Here some of the terraces have wide platforms and steep, angular backslopes. On the southwest sides of the islands, because of their leeward position, the terraces usually have narrow platforms and gentle backslopes. In some cases, especially on the southwest side of Aruba, the backslopes are poorly defined.

Not only are there differences in terrace development between the windward and leeward coasts of Aruba, Bonaire, and Curaçao; the islands do not have the same number of terraces or the same sequence of terrace size. Aruba has three terraces, the second and third being the largest. Bonaire also has three terraces, but the first and third are the most extensive. Curaçao has five terraces, of which the first, third, and fifth are largest.

The individual terraces—first, second, third, etc.—on all three islands are mapped with the same symbol. This similarity of position relative to the sea does *not* imply that the terraces are of the same age. The reader is reminded of this point by the subscript letters in the legend of each map, which indicate that the numbering applies to that island only. A tentative correlation of the terraces with interglacial phases is also indicated by the

[13] *Sailing Directions for the West Indies*, Hydrographic Office Pub. No. 129, U. S. Navy (6th ed.; Washington, D. C.: Government Printing Office, 1949), Vol. 2, p. 260, 264.

[14] This solution weathering of the terrace platform was called to my attention by Mr. de Buisonjé, who pointed out instances where limestone blocks had come to rest on the platform as a result of caving of the sea cliff. Because of the protection against solution weathering provided by the blocks, the platform beneath them stands several feet above the unprotected surface.

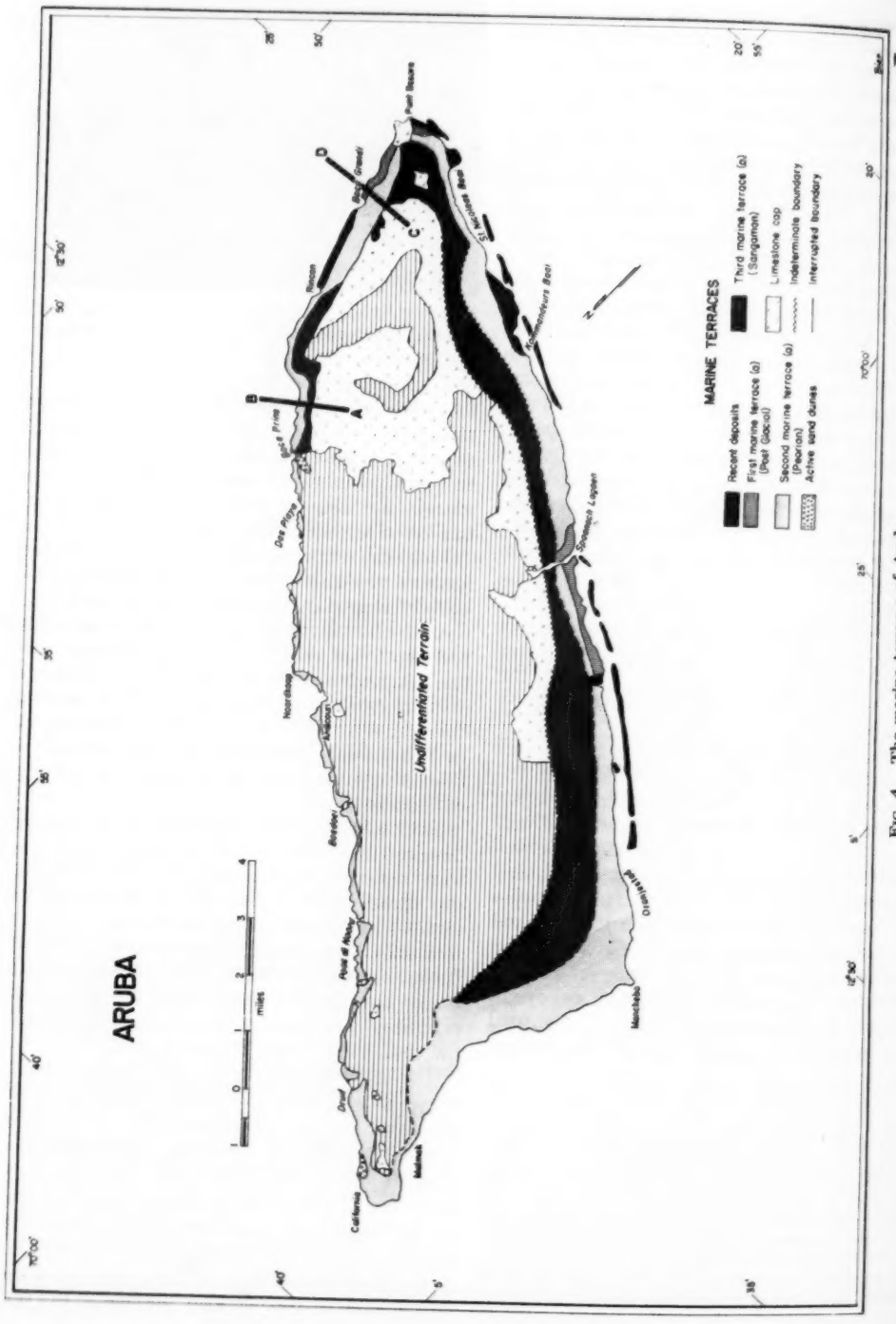


FIG. 4. The marine terraces of Aruba.

names of the interglaciations in parentheses in the map legends. Place names on the maps and in the text follow the spelling on official Dutch maps.

On all three islands the terraces are mainly features of marine abrasion. Their platforms have only a thin (generally 6 feet or less), partial cover of coral limestone that does not appear to extend into the former rough-water zone near the back edge. Only the terraces on the southwest coast of Aruba and on the south coast of Bonaire are chiefly depositional features with platform deposits of undetermined thickness.

Aruba

The terraces on Aruba are vaguely defined. The first (or lowest) terrace occurs as a small step along the east coast between Boca Grandi and Punt Basora and for a short distance along the mid-section of the southwest coast (Fig. 4). The second terrace completely encircles the island, but locally it can scarcely be distinguished from the third, which extends around the southern end and along the southwest side. Only on the southeast coast between Boca Prins and Punt Basora are the terrace platforms bounded by clearly visible backslopes.

The first terrace is so small that it could easily be overlooked in the field.[15] The terrace platform is rarely more than 200 feet wide, and the backslope is seldom over five feet high. The terrace ranges from 24 to 30 feet in elevation at the back edge near the south end of the island, and its surface apparently slopes northward (Fig. 10A).

The second terrace varies in appearance from place to place along the coast. On the southeast end of the island the terrace has a well-defined backslope 15 to 50 feet high (the latter height occurs where the third terrace is absent) and a platform varying in width from 150 to 250 yards between Boca Prins and Punt Basora (Fig. 5) to less than 50 yards between Punt Basora and St. Nicolaas Baai. Following its formation this part of the terrace (from Boca Prins past Punt Basora to St. Nicolaas Baai) was unequally uplifted in such a manner that its surface now slopes northward from Punt Basora (Fig. 10A).

FIG. 5. The second marine terrace near Rincon on Aruba. The wave-cut notch can be seen to the right of the truck.

To the northwest of St. Nicolaas Baai on the west side of the island the backslope of the second terrace quickly decreases in height and disappears. Just before it vanishes, the backslope trends inland, indicating a widening of the terrace platform (Fig. 4). Between St. Nicolaas Baai and Kommandeurs Baai the platforms of both the second and the third terrace are covered with a mantle of alluvial outwash, which completely buries the boundary between them. Northwest from Kommandeurs Baai this detritus is absent, but the inner edge of the second terrace continues to be obscure and was located, in part, by the position of a vague and discontinuous backslope and in part by noting differences in limestone. The deposit believed to belong to the second terrace is a hard coral limestone, and that believed to belong to the third is a soft, limy marl.

Again, the second terrace lacks a clearly defined backslope on the northeast coast between California and Boca Prins (Fig. 4). Here the terrace was cut into crystalline rock, and a thin limestone bed was subsequently deposited on its abrasion surface. Because of the relatively rapid weathering of the crystalline rock, the backslope has become a smooth slope that is frequently indistinguishable from the steep terrain rising above the terrace. Marine erosion has destroyed much of the terrace platform, so that nothing remains except a narrow, but nearly continuous, limestone-capped ledge along the coastal margin.

[15] The width of the first terrace has been exaggerated, as has that of the second between Boca Prins and Rincon and Punt Basora and St. Nicolaas Bay, in order to make them visible on Figure 4.

BONAIRE
MARINE TERRACES
Recent deposits
First marine terrace (b) (Pearian)
Related bay terraces (b)
Second marine terrace (b) (Sangamon)
Third marine terrace (b) (Aftonian and Yarmouth)
Limestone cap
Indeterminate boundary
Undifferentiated Terrain
Undifferentiated Terrain
KLEIN BONAIRE
Lac
Lagun
Lac
Barcadera
Kralendijk
Punt Vierkant
Boca Cruna
Boca Cruna
Boca Cocolishi
Boca Slagbaai
Boca Bartol
Boca Onima
Boca Spelonk
Playa Chikitu
Salina Terra
Salina Gotomeer
Salina Slagbaai
Salina Bartol
Salina Cocolishi
Salina Mathijs
Salina Onima
Salina Spelonk
Salina

The third terrace on Aruba has a clear-cut backslope, about 25 feet high, only on the southeastern end of the island between Boca Prins and Punt Basora. Between St. Nicolaas Baai and Malmok the boundary between the third terrace and the limestone cap could not be clearly established. Here, however, the limestone of the third terrace is soft and that of the cap is hard and massive, a contrast that was used to distinguish the two features.

Bonaire

Because of the large expanse of the first and third terraces on the southern part of the island, nearly half of the surface of Bonaire is covered by terrace platforms (Fig. 6). Nowhere else in the leeward islands of the Netherlands Antilles are terrace platforms so well developed. On the southern end of Bonaire the fore-edge portion of the third terrace platform merges with the back-edge section of the first to form a single, smooth surface, a surface that slopes gently southward and passes beneath the sea. Near Lac on the east coast and Punt Vierkant on the west coast the first terrace backslope between the two platforms commences as a short gentle slope. Northward from Lac and Punt Vierkant the backslope becomes an escarpment about 5 feet high and gradually increases in height in that direction. On the north coast the backslope of the first terrace is about 25 feet high. Concomitant with the increase in height of the backslope, the first and third terraces increase in elevation toward the north, the third increasing at a more rapid rate (Fig. 10C). A little over one-half of the way towards the northwestern end of the island the second terrace makes its appearance on both coasts (Fig. 6).

The first terrace, with its relatively wide platform, completely encircles Bonaire except for a narrow gap near Barcadera. Since its development, the first terrace has been unequally uplifted or arched in several places. The most clearly defined arch occurs on the northeast coast, the apex of the arch lying near Boca Onima (Fig. 10C).

The larger part of Klein Bonaire, west of the main island, is composed of the first terrace platform. The second terrace does not occur on this island, and the third is found only on a small section on its eastern part (Fig. 6). Along its eastern side the higher

FIG. 7. The marine terraces near Boca Onima on Bonaire. From the foreground to the background are the first, second, and third terraces.

surface is separated from the first terrace by a pronounced, though low, backslope. Elsewhere there is no clearly visible boundary between the two terrace surfaces.

The second terrace lies mainly to the west of the narrow waist of Bonaire, extending along both the northern and the southern coasts for about six miles (Fig. 6). This second terrace has been uplifted and arched in the same manner as the first, the apex of its principal arch lying near Boca Onima (Fig. 10C). Along the southern coast the terrace is narrow and poorly defined, but along the northern coast the terrace is, at its western end, a well-defined though fragmented feature. At its eastern end, on the northern coast, the second terrace merges laterally with the first. Here the second terrace can be first observed as a low, narrow ledge gradually appearing along the back edge of the first. Westward this ledge widens, forming a terrace; the elevation of the terrace increases, and the backslope between it and the first terrace becomes progressively higher, reaching a maximum of about 30–40 feet (Fig. 7). Eventually, in the vicinity of Boca Onima, the second terrace is larger than the first (Fig. 14).

The outstanding features of the third terrace are the height of its backslope (60 to 100 feet) between Boca Oliva and Lagoen and the great width of its platform in the central part of the island (Fig. 6). West of Boca Oliva the platform narrows suddenly, and as a result of subaerial erosion becomes a discontinuous series of small, widely separated, limestone-capped mesas. To the south, in the vicinity of Lagoen and Boca Washikemba,

FIG. 8. The marine and bay terraces of Curaçao.

part of the terrace—as well as all of the central limestone cap—has been destroyed by subaerial erosion. Still farther south, where the inner margin of the terrace swings westward, its platform again becomes a broad (up to two miles wide) surface, gently sloping seaward. Here the backslope is absent; the coral platform abuts against low, rolling hills and its inner margin is covered by an outwash of alluvial detritus. The backslope reappears as a low bluff to the north of Kralendijk, where the third terrace is cut into the limestone cap. From Kralendijk to Salina Tam the platform gradually narrows until it is less than 50 yards wide where the terrace ends. Elevation measurements from west of Boca Oliva to Lagoen show that the third terrace has been tilted and probably arched in the same fashion as the first and second (Fig. 10C).

Curaçao

The main terraces between St. Joris Bay and Boca Ascencion on the north coast of Curaçao were eroded by vigorous trade-wind seas, and have wide platforms and high backslopes of coral limestone. Northwestward from Boca Ascencion and around the island to Bullen Bay only the first terrace is continuous (Fig. 8). Higher terraces in the northwest are found only where patches of limestone cap have been preserved. On the lee coast between Bullen Bay and Point Kanon at the southern tip of the island the terraces, even where preserved, have narrow, discontinuous platforms and low backslopes.

The first terrace is best preserved on the windward coast between Point Kanon and Boca Ascencion. Here the terrace platform is generally free of alluvial detritus, and ranges from 400 to 800 yards in width. The backslope is an abrupt escarpment 20 to 30 feet high, and nearly everywhere has a clear wave-cut notch at its base (Fig. 3). After it was cut, this first terrace was unequally uplifted and as a result has been longitudinally arched in several places. Three clearly defined arches occur along the northeast coast: one near St. Joris Bay, another near Klip, and the third approximately midway between Bartool Bay and Boca Grandi (Fig. 11F). Several less clear arches can also be detected along the southwest coast.

The second terrace is small and discontinuous. Its greatest area of uninterrupted occurrence is on the east coast between Hato and Boca Ascencion (Fig. 8). Even here the second-level platform is 250 yards or less in width, and lies only slightly (8–10 feet) below the surface of the third terrace platform (Fig. 9). The second terrace has also been arched during the course of its uplift. Its arching coincides with that of the first. This coincidence is especially clear on the northeast coast between Point Kanon and Boca Ascencion (Fig. 11F).

FIG. 9. The marine terraces near Boca Ascencion on Curaçao. The escarpment in the middle ground (c. 30 feet high) is the backslope of the first terrace, and the escarpment against the sky (c. 60 feet high) is the backslope of the third terrace. The backslope of the second terrace (c. 10 feet high), shown by the arrow, is scarcely visible.

The third terrace also has a discontinuous distribution. Unlike the second, however, the third is a large feature. On the windward coast southeast of Boca Ascencion the terrace platform is from 200 to 400 yards wide, and its old sea cliff is higher than that of the first, being up to 60–70 feet high (Figs. 2 and 9). The platform here is covered by four large indurated fossil dunes of aeolian calcarenite.[16] Like the first and second, the third terrace is arched, its points of highest elevation coinciding closely in position with those of the lowest terraces (Fig. 11F).

The fourth terrace occurs at only one place on the northeast coast (at St. Joris Bay) and at two places on the southwest coast (Fig. 8). These remnants of the terrace are ill defined, having badly eroded platforms and backslopes. Only a small ledge-like remnant with a 6- to 8-foot backslope, north of St. Joris Bay,

[16] Identification by Dr. J. I. S. Zonneveld and Mr. P. H. de Buisonjé of the Geologisch Instituut of Amsterdam during the summer of 1956.

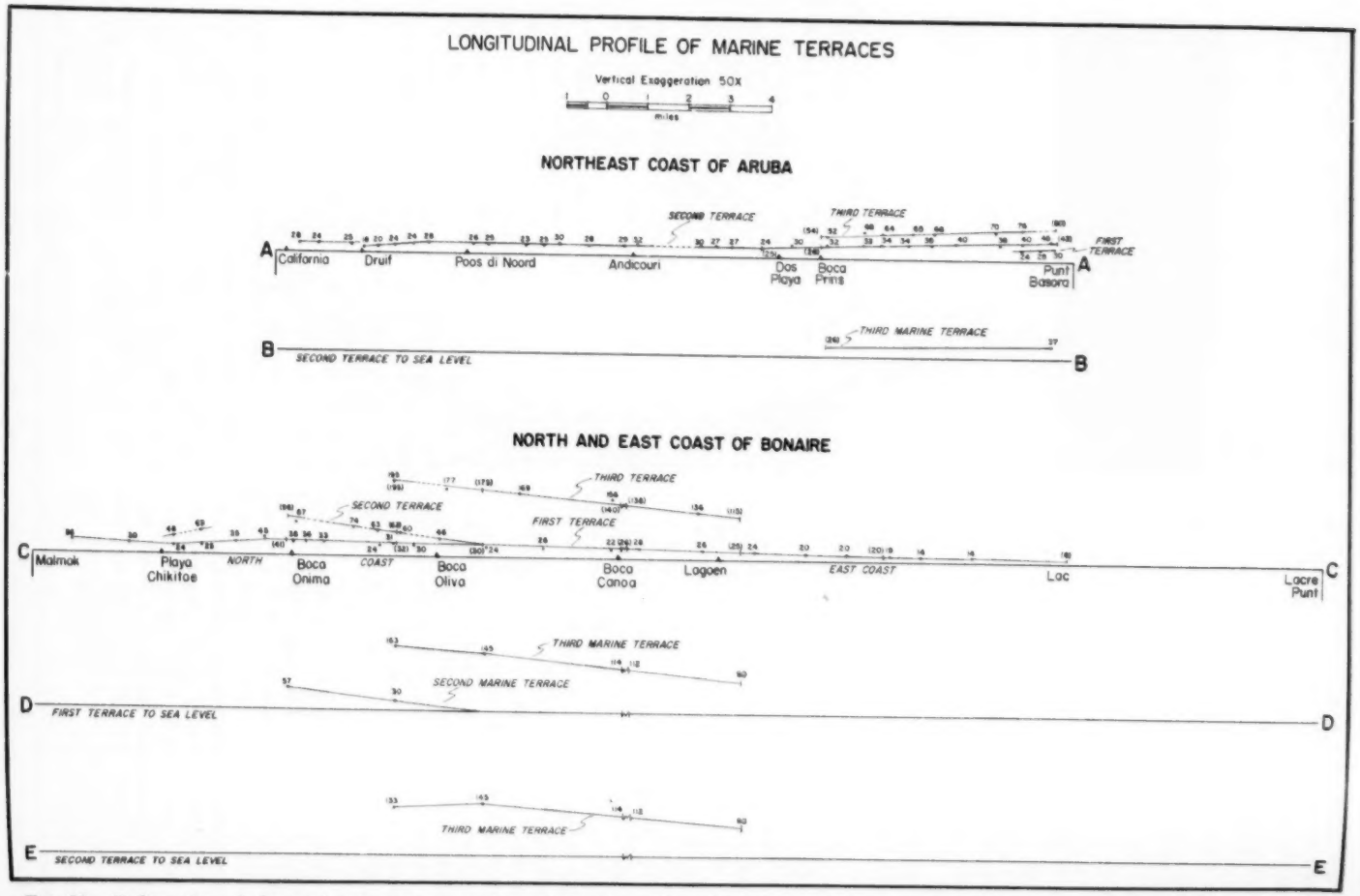


FIG. 10. Deformation of the marine terraces on Aruba and Bonaire. The solid lines in profiles A and C indicate areas of preserved terraces where dependable measurements of elevation were made. The dashed lines show sections where the terraces have been destroyed by marine or subaerial erosion.

FIG. 10. Deformation of the marine terraces on Aruba and Bonaire. The solid lines in profiles A and C indicate areas of preserved terraces where dependable measurements of elevation were made. The dashed lines show sections where the terraces have been destroyed by marine or subaerial erosion.

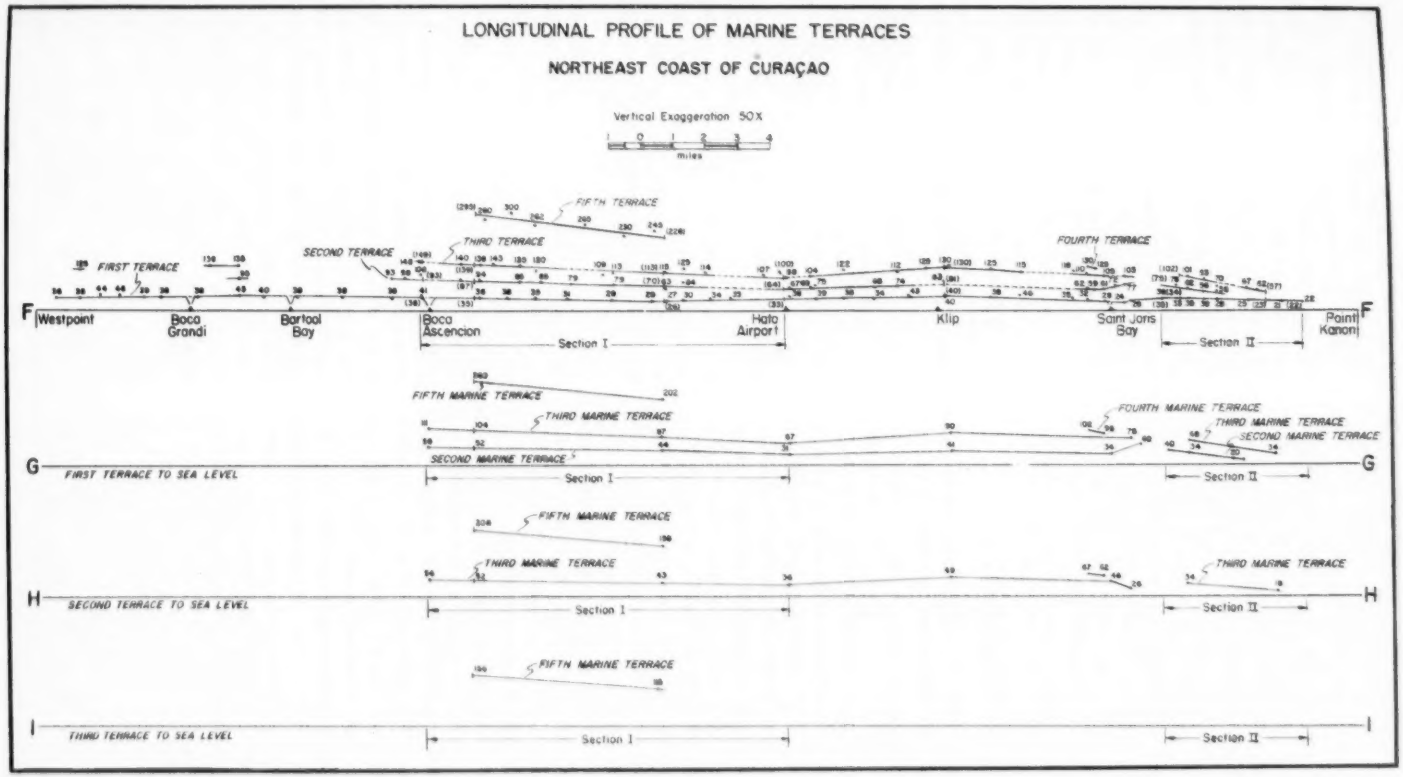


FIG. 11. Deformation of the marine terraces on Curaçao. The lines in profile F have the same meaning as those in A and C, Fig. 10.

is clear enough to permit measurement of its elevation. Because of the scarcity of other fragments no arching can be detected.

The fifth terrace, for the greater part of its extent, is represented only by a discontinuous series of large limestone-capped plateaus situated mainly on the northeast side of the island (Fig. 8). The old sea cliff (30 to 40 feet high) occurs only at the waist of the island, and has elsewhere apparently been destroyed by subaerial erosion. Elevation measurements made here at the base of the old sea cliff show that the terrace is tilted in the same fashion as the lower terraces in the vicinity (Fig. 11F).

BAYS AND BAY TERRACES

Pouch-shaped bays are distinctive features of the coastal landscape on Bonaire and Curaçao. The pouch-like shape results from the drowning of valleys eroded into both the interior crystalline rocks and the rim of terrace and cap limestone. On the crystalline interior the valleys are open and rounded in cross-section, but in the limestone the valleys have steep slopes and resemble straight slots cut through the rim. When they were drowned, the flaring valleys in the crystalline region were transformed into wide bays, while the slot through the limestone rim became a narrow passage. At least two such bays formerly existed on Aruba also (Druif and Spaansch Lagoen, Fig. 4), but the interior portions have been filled with alluvium.

A single small alluvial terrace, a remnant of a former bay fill, occurs around the margins of many of these bays on Bonaire and Curaçao (Figs. 6 and 8). There is also a trace of a similar terrace on the east side of Spaansch Lagoen on Aruba. On all three islands this terrace is composed of a sandy marl that contains numerous shells and occasional bits of coral. The terrace material was deposited during a period when sea level was relatively higher than at present, and the bays were being completely or partly filled with alluvial detritus washed from the land and with the remains of marine organisms then living in the bays. At some time following its deposition the fill was exposed to, and very nearly destroyed by, subaerial erosion, the remnants forming a bay terrace. To the extent that this bay terrace can be related to the marine terraces, it constitutes additional significant evidence concerning the origin of the individual marine terraces of the islands.

The bay terraces on Bonaire and Curaçao can be definitely related to the first marine terrace. On the west side of Goto on Bonaire the surface of the bay terrace extends well into the slot (Fig. 6), and hand-level sights indicate that it has the same elevation as the marine terrace. Hand-level sights also show that the surface of the bay terrace in Slagbaai and the surface of the first marine terrace have similar elevations. The bay terrace on Bonaire, then, is a depositional equivalent of the first terrace. A similar conclusion can be made concerning the bay terrace on Curaçao. The terrace of Spanish Water Bay can be followed on the sides of the bay entrance to points where it merges with the first marine terrace (Fig. 8). The remnants of the terrace in St. Joris Bay can be correlated with the first marine terrace by means of hand-level sights. The terraces in Schottegat, in Piscadera Bay, and the tiny remnants near the mouth of St. Martha Bay all have much the same elevation as the first marine terrace.

On Aruba the bay terrace consists only of scattered patches of marl along the eastern slopes of Spaansch Lagoen. The elevation of these fragments could not be accurately measured, though they appeared to extend at least 10 to 15 feet above the surface of the lagoon. This fact, together with the circumstance that the marl could not be traced to the coast, made it impossible to relate the fragments to any one marine terrace.

DEFORMATION OF TERRACES

To show the nature and amount of deformation or arching of the terraces, their elevations are plotted in longitudinal profile on figures 10A, 10C, and 11F. Straight lines indicating the present general position of the terraces are fitted to the plotted points. The plotted points show local variations in terrace elevation that may be the result of small local variations in uplift, subtle differences in local erosion or deposition, or errors in measurement. There is no means of distinguishing among these causes or of making allowance for them; consequently a line drawn from point to point would give an erroneous impression of the uplift as well as an illusion of greater accuracy than the measurements possess. Straight lines are easier to use in the

analysis of the profiles undertaken here. Moreover, at the time of origin the back edge of a marine terrace lies in a horizontal line. The straight lines are drawn to connect the greatest number of points, and where the points are scattered the lines are drawn in median positions. The lines on the profiles, then, do not everywhere indicate the precise elevations of the terraces but represent generalizations of their present altitudes.

Although longitudinal profiles were made of the terraces in all sections of the islands, only those from the windward coasts are illustrated because of the better preservation of the terraces in those areas. The profiles from the leeward sides of the islands demonstrate the same kind of deformation as those on the windward sides, but the evidence is less complete.

From inspection of the profiles it is evident that the terraces are not the sole result of wave action occurring at different levels during high stages of interglacial seas. The bending or arching of the terraces on all three islands is a clear indication of the instability of the land. Further, the profiles show that the terraces were not all deformed together. Their platforms are not parallel, but are inclined, or tilted, by different amounts. Deformation of the terraces appears to have occurred throughout the period of terrace cutting.

To determine the progress of arching of the terraces, and thus of the general uplift of the islands, the profiles are reconstructed so that the back edge of each terrace, commencing with the lowest, is horizontal at sea level. This reconstruction is effected by subtracting the high and low elevations of the first marine terrace *line* (largely absent on Aruba, where the second is used) from those of each of the higher terrace lines. The results of this subtraction, when plotted on the profile, approximate the position of the higher terraces at the time the first was being formed. The process is then repeated with the second terrace, etc. (Elevations of the terrace *lines*, in contrast to terrace elevations, are in parentheses on Figs. 10 and 11.) This procedure gives an opportunity for calculating the amount of tilt in feet per mile (to the nearest tenth) that the terrace immediately above sea level on each profile had already experienced at the time of origin of the next lower one. Such calculations of tilt were made for the terraces on selected portions of the windward coasts of all three islands. The tilt of the first terrace can, of course, be estimated by reference to present sea level. Those sections selected are within well-defined regions of localized arching.

Aruba. The first terrace is of such limited extent that the amount of its tilt cannot be evaluated. The second therefore is treated as the lowest. Between Dos Playa and Punt Basora, 7 miles, it ranges in elevation from 25 to 43 feet, and so has an inclination to the north of 2.6 feet per mile (Fig. 10A). When the second terrace is reduced to sea level (Fig. 10B), the third terrace is found to range in elevation from 26 to 37 feet in 5 miles, thus having a northward slope of 2.2 feet per mile.

Bonaire. On the north coast, between Boca Canoa and Boca Onima, the first terrace has an elevation between 26 and 41 feet: a difference of 15 feet in 8 miles, or a slope to the east of 1.9 feet per mile (Fig. 10C). When the first terrace is brought to sea level by the method described earlier (Fig. 10D), the second terrace shows a slope to the east of 11.6 feet per mile. With the reduction of the second terrace to sea level (Fig. 10E), the third terrace has an eastward tilt of 8.8 feet per mile.

Curaçao. Only the parts of Curaçao's windward coast between Boca Ascencion and Hato Airport (section I) and between St. Joris Bay and Point Kanon (section II) are considered here, because these are the areas in which the terraces form the most complete sequence and are the best preserved. The terraces in the intervening section, between Hato Airport and St. Joris Bay, cannot be satisfactorily evaluated because of the infrequent occurrence of the second terrace and the presence of extensive alluvial fans on the platform of the first terrace.

The first terrace in section I lies between elevations of 26 and 38 feet within a distance of 7 miles, and so has a tilt to the southeast of 1.7 feet per mile (Fig. 11F). When the first is brought to sea level, the second terrace is found to lie between elevations of 31 and 55 feet, within a distance of 11 miles, showing an inclination of 2.2 feet per mile (Fig. 11G). When the second terrace is brought to sea level, the third terrace lies between elevations

of 36 and 56 feet in 11 miles having a southeastward inclination of 1.8 feet per mile (Fig. 11H). When the third is brought to sea level, the six-mile section of the fifth lies between elevations of 115 and 156 feet and has an easterly tilt of 6.8 feet per mile (Fig. 11I). When the same procedure is applied to section II, the tilt of the first terrace is seen to be 2.9 feet per mile, of the second 8 feet per mile, and of the third 5.3 feet per mile (Fig. 11F–I).

For easy comparison of the deformation of the terraces on the three islands, the amounts by which the terraces have been tilted between successive periods of terrace cutting are listed in tabular form below:

Aruba		*Bonaire* (north coast)	
Third terrace	2.2 ft./mile	Third terrace	8.8 ft./mile
Second terrace	2.6 ft./mile	Second terrace	11.6 ft./mile
First terrace	Not available	First terrace	1.9 ft./mile

Curaçao

Section I		*Section II*	
Fifth terrace	6.8 ft./mile	Fifth terrace	Not available
Third terrace	1.8 ft./mile	Third terrace	5.3 ft./mile
Second terrace	2.2 ft./mile	Second terrace	8.0 ft./mile
First terrace	1.7 ft./mile	First terrace	2.9 ft./mile

From this study of the profiles it becomes evident that the deformation or arching of the terraces was a repeated process. On all three islands each higher terrace was arched before the emergence of the next lower one from the sea. Since the arching is only a local modification of the general uplift of the islands, it is evident that the islands have been subjected to repeated movement. In view of this fact, the obvious explanation of the terraces is that they were eroded during the still-stands in the uplift. This conclusion presumes, however, a stable or nearly stable sea level and relatively rapid and sudden uplift of the land to overcome wave erosion, which otherwise might be capable of maintaining a platform against the uplift. These presumptions would lead to the conclusion that some or all of the terraces are postglacial in age.

This conclusion is untenable. The first terrace on Bonaire and Curaçao was eroded before the last major glacial lowering of the sea level, as follows from the existence of the depositional bay terrace on the two islands. The deformation of marine terraces suggests that the only movement of the islands has been upward. If this is the case, the bays or drowned stream valleys and their alluvial terraces cannot be related to a rise and fall of the islands. They must result from rises and falls in sea level. The flooding of the lower parts of the valleys that created the bays would result from the last deglacial rise in sea level. The prior erosion of the valleys into the bay fill that left remnants as an alluvial terrace occurred during the lowered sea level that accompanied the last glaciation. The deposition of the alluvial bay fill obviously required flooding of the valleys, which resulted from the last major interglacial high of sea level. The erosion of the valleys prior to the deposition of the bay fill probably occurred during the lowered sea level of the next-to-last glacial age. Since the first marine terrace on Bonaire and Curaçao can be correlated with the bay terrace, the former must also have been cut during the last interglacial high sea level.

The higher terraces also probably resulted from marine erosion during earlier interglacial high sea levels, since glacially caused rises and falls in sea level occurred several times during the Pleistocene. There is, then, no necessity to assume repeated or spasmodic uplift of the land in an explanation of the terraces on Aruba, Bonaire, and Curaçao. The uplift may also have been continuous; and the cliffs, backslopes, or terraces at different elevations on the islands may simply be an indication of successive high-level stands of interglacial seas.

The idea of continuous and slow uplift and deformation of the land is not new. The assumption of continuous, slow uplift is an important part of W. Penck's system of landform analysis.[17] Indeed, evidence for continuous movement has been observed at widely separated places. According to Fisk, the vertical separation of the Mississippi Valley alluvial terraces was caused by continuous and slow uplift resulting from isostatic adjustment to the immense burden of alluvium deposited in the Mississippi delta throughout the Pleistocene.[18] Continuous, slow movements have been reported from Southern California, Italy,

[17] Walther Penck, *Morphological Analysis of Land Forms*, trans. by H. Czech and K. C. Boswell (New York, 1953), pp. 1–18.

[18] H. N. Fisk, *Geological Investigation of the Alluvial Valley of the Lower Mississippi River* (Vicksburg: Miss. River Comm., 1944), pp. 57–69.

and Japan, all regions of tectonic instability.[19] The long-time operation of these continuous movements resulted in significant changes in the elevation of the land.

It should not be surprising that the three islands have also undergone continuous, though slow, uplift throughout the Pleistocene. Aruba, Bonaire, and Curaçao are situated on or near the axis of the Aruba–Orchilla geanticline, a part of the West Indian structural zone where great crustal deformation lasted to the late Tertiary.[20]

Steady, though slow, deformation of the islands is, then, more plausible than a few large uplifts widely spaced in time. If continuous uplift is not an acceptable interpretation, then the evidence must be construed to mean that deformation of the terraces occurred during periods of glacial low sea level, since marine erosion during periods of high seas could well eliminate all traces of uplift. The depression of the North American continent during periods of glaciation is widely accepted, but a compensating rise in the Caribbean has not yet been suggested. A sizable stumbling block hampers such a suggestion. The Gulf geosyncline lies between the two areas, and probably would have prevented a compensating rise to the south of it. The assumption of uplift of the islands during glaciation and stability during interglaciation is highly questionable; it is much less plausible than the postulate of a continued slow rise of the islands during the Pleistocene.

Nothing can be said about either the rate or variations in the rate of rise of the land because of the lack of information concerning the absolute time lapse between periods of terrace cutting. The great difference in elevation between the second and third terraces on Bonaire and the third and fifth on Curaçao can mean a greater rate of uplift during the intervening periods of terrace cutting, or it can mean a longer span between the respective stands of high sea level. The local evidence is insufficient to permit a decision between these possibilities.

Because of its evident instability, the land cannot serve as a stable datum; and the question arises as to the reliability of the sea as datum. Here uncertainties arise because the degree of deglaciation in each interglacial period is not known. The possibility of deformation of the ocean basins must also be considered. Deformation of the basins occurred in such a manner as either to increase or decrease their volume and thereby to produce either a lowering or a rise of sea level. Because of these uncertainties the levels reached by different interglacial high seas remain problematic.

However, a widely held theory of terrace formation postulates an increase in the size of the ocean basins through downwarping, which caused a general lowering of the sea level during the Pleistocene. In this interpretation, the land is held to be stable and the sequence of terraces from higher and older to lower and younger is ascribed solely to the lowering of sea level. Areas in which this theory has been widely applied are the Mediterranean, the islands in the Pacific, and the Atlantic coast of the United States.[21] A notable similarity of the Mediterranean and the Pacific regions is that they are both areas of marked instability. The Mediterranean is well known for its earthquakes, and recent bathymetric measurements in the Pacific basin show that its floor has been bent and faulted; volcanic activity around the rim and within the basin is one of its most notable features. The central Atlantic coast lies between the delta of the Mississippi and New England. It has been proved the Mississippi Delta sank throughout the Pleistocene,[22] while New England has had considerable uplift since the last glaciations.[23] It is difficult to imagine that the area in between remained stable. It is also difficult to defend the assumption that the land remained stable while eustatic variation in sea level produced terraces.

Since the deformation of the terraces suggests that the islands are unstable and have been rising continuously, it is necessary to

[19] James Gilluly, "Distribution of Mountain Building in Geologic Time," *Bulletin*, Geological Society of America, Vol. 60:1 (1949), pp. 561–90; James Gilluly, A. C. Waters, and A. O. Woodford, *Principles of Geology* (2d ed; San Francisco, 1959), pp. 112–14.

[20] H. H. Hess, "Gravity Anomalies and Island Arc Structure with Particular Reference to the West Indies," *Proceedings of the Philosophical Society*, Vol. 79 (1938), pp. 71–96; "Geological Reconnaissance of the Island of Margarita." *Bulletin*, Geological Society of America, Vol. 60:2 (1949), pp. 1,857–68.

[21] Zeuner, *op. cit.*, pp. 283–307.

[22] Fisk, *op. cit.*, pp. 57–69.

[23] R. F. Flint, *Glacial and Pleistocene Geology* (New York, 1957), pp. 247–55.

consider the effect of eustatic changes in sea level on slowly and continuously rising land. The rise could occur at a rate less than or greater than the rate of marine erosion (though unlikely, uplift equal to marine erosion would have the same result as uplift occurring at a rate less than that of marine erosion). Because the arching or tilting of the terrace surfaces is slight, and the total uplift since the onset of terrace erosion is small (136 to 195 feet on Bonaire, 230 to 300 feet on Curaçao), it is probable that the uplift of the islands occurred at a rate less than that of marine erosion.

Under this condition the submerged marine platform would be lowered and extended landward by marine erosion during relatively stable, interglacial high sea level (Fig. 12A). This process would continue until sea level was depressed during the subsequent glacial epoch. During the lowering of the sea a coastal plain would gradually be exposed until the sea reached its lowest position. At this low, stable level, marine erosion would again wear down and broaden the submerged platform, developing a cliff against the slowly rising land (Fig. 12B). With the next interglacial rise of sea level the cliff would advance landward as marine erosion extended the submarine platform across and somewhat below the surface of the earlier coastal plain. At the conclusion of this eustatic cycle the former marine platform and sea cliff would be exposed as a marine terrace (Fig. 12C).

Eustatic changes in sea level could also erode terraces if the rate of uplift of the islands were greater than the rate of marine erosion, though this rate of uplift seems unlikely. During relatively stable interglacial high and glacial low sea levels the rise of the land would expose a widening coastal plain. This widening would be accelerated during the intervening glacial subsidence of sea level. The next interglacial rise of the sea would at first cancel the effect of uplift and then, if the rate of rise exceeded uplift, cause the sea to advance across the coastal plain eroding a sea cliff as it undercut the surface of the plain. When uplift again exceeded the rate of marine erosion a new and lower coastal plain would be exposed at the foot of the sea cliff cut during the rising stage of sea level.

The terraces of Aruba, Bonaire, and Curaçao can be explained as having been formed on a constantly rising land during periods of changing interglacial sea levels.

Though the absolute levels attained by different interglacial high seas are difficult to determine, the lower two terraces of Bonaire suggest that there was little difference in height among the seas responsible for their erosion, if the local evidence of continuous rise of the islands is accepted as conclusive. On the south side of the island the first terrace has an elevation of 6 feet. This fact, taken in conjunction with the evidence of slight uplift at the same end of the island, indicate that the interglacial high sea that eroded the terrace was nearly at the present level of the sea. On the north coast of Bonaire the second terrace merges laterally with the first, showing that the interglacial high sea levels that eroded the first and second terraces also came to nearly the same level. Uplift, rather than a general lowering of sea level, appears to be the main reason for the superposition of the marine terraces on the three islands.

CORRELATION OF THE TERRACES AMONG THE ISLANDS

If the sea stood at high levels through long periods of time, it would cut large, contemporaneous terraces on all three islands. These terraces can be correlated on the basis of size if certain variable factors can be equalized. The size of a cut terrace involves not only the width of its platform but also the height of its backslope (Fig. 13). Terraces of comparable magnitude are eroded during the same period of time if they have the same exposure to the sea and the same underlying material, and if they are cut into terrain having the same general seaward slope.[24] Fortunately, on Aruba, Bonaire, and Curaçao these factors are relatively constant. Exposure to the sea (northeast coast) is the same, underlying material (diabase, Cretaceous sedimentary rocks, and Pleistocene limestone) is the same, and the general seaward slope of the terrain is similar in most places. The important variable factor influencing the magnitude of ter-

[24] W. S. T. Smith, "A Topographic Study of the Islands of Southern California," *University of California Publications in Geology*, Vol. 2, No. 7 (1900), pp. 205–30.

FIG. 12. Formation of terraces during eustatic changes in sea level when the rate of marine erosion exceeds the rate of rise of the land.

A: Sea level remains stable at a high level during one unit of time.
   a. Lateral marine erosion accomplished in unit time.
   b. Amount of uplift in unit time.
   c. Amount of downward marine erosion required to compensate for uplift *b*.

   During unit time, cliff ABCD becomes cliff EFC_1D_1 rather than $A_1B_1C_1D_1$ because uplift *b* lies within the competence of downward marine erosion. Following the formation of cliff EFC_1D_1 glacial lowering of sea level causes the shoreline to recede to IJ.

B: Sea level remains stable at a low level during one unit of time.
   For simplicity it is assumed that the sea subsides so rapidly to its stable low position that no uplift occurs. The position of the land at the conclusion of the subsidence is shown by the dashed line.
   a. Lateral erosion in unit time.
   b. Amount of uplift in unit time.
   c. Amount of downward erosion required to compensate for uplift *b*.
   d. Amount of downward erosion required to compensate for uplift *b* and lateral erosion *a*.

C: Sea level rises to a stable high position, rising faster in one unit of time than the rate of uplift. The rise follows the two unit periods of stable low sea level that resulted in sea cliff OPQR.
   a. Lateral erosion in unit time.
   At the end of unit of time sea cliff OPQR has retreated to STUV and is reduced in height by the amount by which the rise of the sea exceeds the uplift of the land. Because it was eroded by a rising sea, the newly formed platform of marine erosion UVQR has a steeper slope than the platform seaward from QR.

   Eustatic fall and rise of sea level produces terrace EFCDST on a coast uplifted at a rate less than the rate of marine erosion.

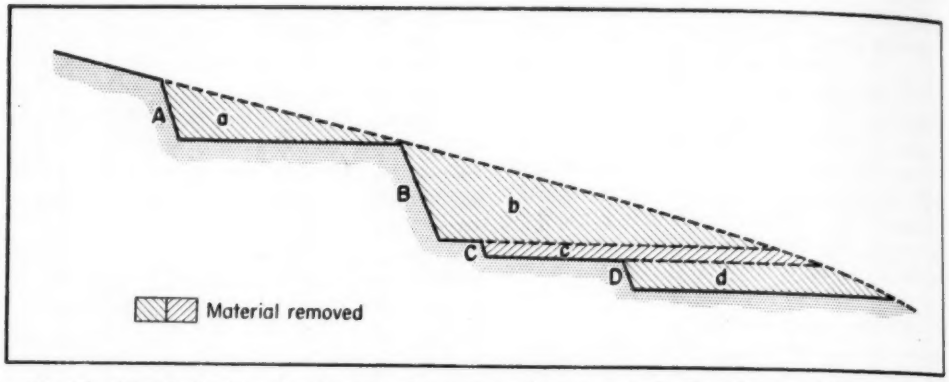


FIG. 13. Magnitude of marine terraces. A sequence of terraces, A, B, C, and D, eroded into terrain of nearly uniform slope and underlain by uniform material. Though terraces A, C, and D have wider platforms than B, terrace B requires the removal of more material than A, C, and D. Underlying material and wave energy being equal, more time is required to remove *b* than *a*, *c*, or *d*.

races on the islands is the duration of marine erosion.

In order better to compare the magnitudes of terraces among the islands, transverse profiles of the terraces on the northeast coast have been drawn for each island. These show representative examples of width of platform and height of backslope on the several islands (Fig. 14).

From examination of these profiles it is evident that the first terrace on Bonaire and Curaçao and the second on Aruba are of comparable magnitude. The first terrace on Aruba and the second on Curaçao are peculiar to those islands. The third terrace on Aruba, though relatively small in comparison with the second on Bonaire and the third on Curaçao, is tentatively correlated with those terraces. Because the fourth terrace on Curaçao is of limited occurrence and poorly preserved, it is difficult to make an accurate assessment of its size but it appears to be small and to have no counterparts on the other islands.

The third terrace on Bonaire is substantially larger than the fifth on Curaçao, both in width of platform and height of backslope. Since each is the highest local terrace, each is presumably the first to have been eroded on its island. However, the third terrace on Bonaire, because of its great size, may represent not only a period of erosion contemporaneous with Curaçao's fifth terrace but also the erosion of an earlier period of high sea level. That these two phases of high-standing sea were not recorded by separate terraces on Bonaire can be explained by assuming relatively slow initial uplift on that island. For easy comparison the tentative correlation of the terraces among the islands is shown in a later tabulation.

TENTATIVE CORRELATION OF THE MARINE TERRACES WITH INTERGLACIAL PHASES

Attempts to relate marine terraces to specific interglaciations are hampered by unknown factors on both sides of the correlation. In a sequence of eustatic marine terraces, one or perhaps more may be missing as the result of destruction by marine erosion during the formation of a later terrace. Terraces may also be missing from the sequence because of a late onset of uplift or its premature cessation. Problems are also encountered when considering the interglacial phases. There have been three, possibly four, major interglaciations with several interstadials or subinterglaciations that divided the glaciations into two or more phases.[25] In addition to the problem of the number of interglaciations, some uncertainty exists as to their duration.[26] In the face of these obstacles, the correlation of marine terraces with interglacial phases attempted here is at best tentative and is introduced mainly to provide material for future discussion.

[25] Zeuner, *op. cit.*, pp. 23–80.
[26] Flint, *op. cit.*, pp. 289–301.

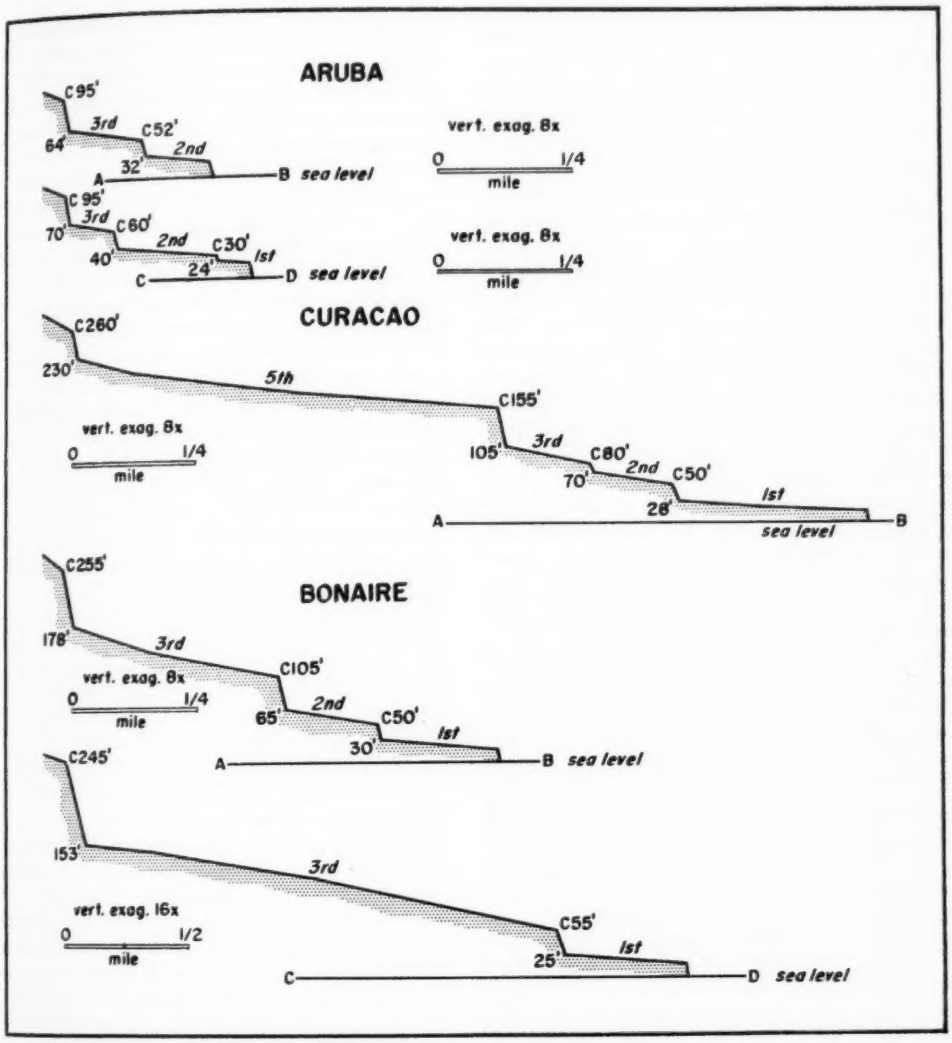


FIG. 14. Profiles of the marine terraces on the northeast coasts of Aruba, Bonaire, and Curaçao.

Since both the number of terraces and the number of interglacials are problematic, their direct correlation by relating the lowest terrace to the last interglacial and each higher terrace to the next older interglacial has obvious weaknesses. The use of differences in the degree of weathering or soil formation on the terrace platforms of the islands as a means of determining their relative age is not practical. The soils on the terraces are thin and scanty; and all that is observed on successively higher platforms is a slight increase in redness of soil. The most satisfactory correlation at present, despite the uncertainty involved, is that based upon magnitude of the terraces and the estimated relative lengths of the interglaciations.

While reliable methods of determining the length of interglacial stages are not yet available, estimates of their relative duration are similar. Kay, basing his calculations on the assumption that the youngest drift sheet is

25,000 years old, measured the thickness of the interglacial leached zone in the drift sheets of Iowa and figured that the Post-Iowan, Sangamon, Yarmouth, and Aftonian stages were, respectively, of 55,000, 120,000, 330,000 and 200,000 years' duration.[27] Revisions in the age of the youngest drift sheet based on Carbon 14 dates shorten the time of the interglacials but do not change their relative duration.

Working on entirely different deposits, Fisk derives estimates similar to those made by Kay.[28] From his study of the alluvial valley of the lower Mississippi and its associated stream terraces, Fisk calculates the duration of the interglacial phases to be about 25,000 years for the Peorian, 140,000 years for the Sangamon, 320,000 years for the Yarmouth, and 260,000 years for the Aftonian.

Basing his estimates on a combination of geologic evidence and the astronomical time scale derived from radiation curves, Zeuner proposes the following tentative time scale: the last interglacial (Sangamon) 60,000 years, divided into two periods of high sea level of unequal duration by a period of increased cold; and penultimate interglaciation (Yarmouth), 190,000 years; and the antepenultimate interglaciation (Aftonian), 60,000 years.[29] When these several estimates are summarized in tabular form it is apparent that the *relative duration* of interglacial high seas is roughly comparable:

Interglacial	*Kay's estimate*	*Fisk's estimate*	*Zeuner's estimate*
Peorian[30]	55,000	25,000	c. 15–20,000
Sangamon	120,000	140,000	c. 30–35,000
Yarmouth	330,000	320,000	190,000
Aftonian	200,000	260,000	60,000

Recent estimates by Emiliani of the duration of the Sangamon, Yarmouth, and Aftonian are much shorter than those given above.[31] Analysis of Pelagic foraminifera from deep-sea cores indicates periodic temperature changes of the surface water in tropical seas. By extrapolating rates of sedimentation based on radio-carbon dates, Emiliani calculates the duration of the interglacial stages as 28,000 years for the Sangamon, 82,000 years for the Yarmouth, and 80,000 years for the Aftonian. Since the extrapolation is based on a short time span (c. 30,000 years), any initial error is greatly exaggerated by the multiple extensions of that time span to encompass the Pleistocene. The duration of the Aftonian in particular may be badly distorted. For this reason the relative duration of the interglacial stages apparent from the tabulation above is to be preferred to that suggested by Emiliani.

All other things being equal, the Yarmouth interglacial high sea should have eroded the largest terrace in the sequences on the three islands. The next in magnitude should be the Aftonian terrace, followed by that formed during the Sangamon. Since the Peorian interglacial appears to be one-half to one-fifth the length of the Sangamon, the Peorian terrace should be relatively small.

The correlation of the terraces with specific interglacial high sea levels proposed here has as its main key the largest terraces on the islands: the fifth on Curaçao and the third on Bonaire. The fifth terrace on Curaçao, the largest of its sequence, should, on the basis of magnitude, date from the Yarmouth interglacial. The third terrace on Bonaire, substantially larger than Curaçao's fifth, probably required the marine erosion available during both the Yarmouth and the earlier Aftonian interglacials for its formation. The third terrace on Aruba and Curaçao and the second on Bonaire are next in magnitude, even though their platforms are relatively narrow because of undercutting during the erosion of the next lower terrace. Their high backslopes indicate the removal of a large volume of material during their formation. In comparison with the Yarmouth fifth terrace on Curaçao, these terraces are of the right order of magnitude for the Sangamon interglacial. This interpretation leaves the second terrace on Aruba and the first on Bonaire and Curaçao to be assigned to the Peorian interglacial, a reasonably satisfactory correlation, since these terraces are smaller than those related to the Sangamon. The first terrace on Aruba may be the result of a postglacial high stand of the sea. The absence of a similar feature

---

[27] G. F. Kay, "Classification and Duration of the Pleistocene Period," *Bulletin*, Geological Society of America, Vol. 42 (1931), pp. 425–466.

[28] Fisk, *op. cit.*, Fig. 75.

[29] Frederick E. Zeuner, *Dating the Past* (London, 1958), p. 145.

[30] Kay's Post-Iowan, Zeuner's Late Monastirian.

[31] Cesare Emiliani, "Pleistocene Temperatures," *Journal of Geology*, Vol. 63, No. 6 (1955), pp. 538–78.

on Bonaire and Curaçao may be the result of its destruction by marine erosion.

If the correlation of the cutting of Curaçao's fifth terrace and the final modification of Bonaire's third with the Yarmouth interglacial is correct, the greater number of terraces on Curaçao can be accounted for. Curaçao, with an elevation of 230–300 feet for its highest terrace in comparison with 136–195 feet for Bonaire's highest, must have been uplifted at a faster rate. Faster uplift would have caused the vertical separation on Curaçao of terraces related to relatively small changes in sea level that probably accompanied the waxing or waning phases of the major rises and falls in sea level. For example, Curaçao's second terrace may be the result of a late, minor fall and rise in the Sangamon high sea level responsible for the third. During the period of minor fall the relatively rapid uplift of Curaçao may have raised the exposed third terrace platform enough so that the subsequent rising sea would erode a new, somewhat lower platform forming the second terrace. In the same manner the fourth terrace may also be the result of a late, minor fall and rise of the sea level that cut the fifth terrace.

The tentative dating of the marine terraces from the three islands is summarized in the following tabulation:

Aruba	*Bonaire*	*Curaçao*
Not present	Third Terrace (Aftonian and Yarmouth)	Fifth Terrace (Yarmouth)
Not present	Not present	Fourth Terrace (Late Yarmouth)
Third Terrace (Sangamon)	Second Terrace (Sangamon)	Third Terrace (Sangamon)
Not present	Not present	Second Terrace (Late Sangamon)
Second Terrace (Peorian)	First Terrace (Peorian)	First Terrace (Peorian)
First Terrace (Post-glacial)	Not present	Not present

If the terraces are correctly dated, then the onset of uplift on Aruba must have been much later than on Bonaire and Curaçao. This conclusion may not be as anomalous as it seems. On none of the three islands did uplift begin at the same time; rather, it apparently began first on Bonaire in the east and then started successively later on Curaçao and Aruba to the west. On Bonaire the uplift began prior to the Aftonian, since the third terrace was cut into the flanks of the island's cap of limestone, a deposit mainly of marine origin. On Curaçao uplift began later, probably prior to the Yarmouth, also because the fifth (and highest) terrace was eroded into the flanks of the island's marine limestone cap. Uplift on Aruba must have begun at some time before the erosion of the Sangamon terrace, since there is no higher terrace on the island. The small first terrace peculiar to Aruba may be evidence of a continuation of fairly vigorous uplift of that island into recent time.

SIMILARITIES TO THE TERRACES STUDIED ON THE CALIFORNIA COAST

The terraces on the Windward Islands of the Netherlands Antilles, and especially those on Bonaire and Curaçao, resemble to an appreciable degree those I studied on the north coast of Monterey Bay, California.[32] The most significant resemblances are the similar deformation (arching) of the marine terraces in both areas and the presence in both places of alluvial terraces that can be correlated with the marine terraces. Thus the evidence from these widely separated localities seems in both cases to support the thesis of origin of Pleistocene marine terraces by simultaneous change in sea level and rise of the land.

[32] C. S. Alexander, "The Marine and Stream Terraces of the Capitola–Watsonville Area," *University of California Publications in Geography*, Vol. 10, No. 1 (1953), pp. 1–44.

CENTRIFUGAL–CENTRIPETAL FORCES IN INTERNATIONAL PETROLEUM COMPANY REGIONS[1]

ROBERT B. McNEE

The City College, New York

SOCIETY consists of organizations as well as individuals; individuals are grouped in combinations, both large and small: churches, families, trade unions, business enterprises, states, and other institutional groups. In the twentieth century, such organizations have multiplied in number, increased in size and complexity, and greatly extended themselves in geographic range. Each of these organizations has a geography; each is an attempt to organize space on the surface of the earth for a particular purpose or cluster of purposes; each seeks to create a special type of homogeneity within a particular area in spite of a general heterogeneity within the area and for the earth as a whole. The geographic study of such phenomena may be called the geography of institutions or the geography of organizations; the geographic study of business systems is the functional geography of the firm.[2]

Although the ultimate purposes of these space-organizing institutions are different and indeed often in conflict, there are many similarities in the processes and problems of spatial organization. Consequently, the functional geography of each system may be presumed to have something in common with the functional geography of all other such organizational systems. In short, a large corporation or labor union is somewhat analogous to an "empire." Although such analogies are very commonly used, few scholars have attempted to probe the actual similarities and differences between these space-organizing institutions. If industrial business organizations and territorial states are, in fact, analogous, the methods and analytical tools of political geography should be applicable, with appropriate modifications, in the study of the functional geography of the firm. Is such the case? For the purposes of this paper, it is assumed that the analogy is sufficiently valid to warrant the examination of appropriate concepts from political geography and to apply them to international petroleum company regions. Then, as a more detailed example, the Standard Oil Company of New Jersey is analyzed in terms of the centrifugal-centripetal concept of Hartshorne.[3]

APPLICABLE CONCEPTS FROM POLITICAL GEOGRAPHY

In the methodological literature of political geography, three concepts seem peculiarly applicable to the analysis of company regions: Gottmann's circulation and iconography,[4] the unified field theory of Jones,[5] and Hartshorne's concept. Since the last seems the most immediately useful, it will receive the most detailed treatment in this paper.

Circulation and Iconography

Gottmann's circulation–iconography concept seems most applicable to the understanding of *international* company regions, especially those most affected by economic nationalism, as in the petroleum industry. Gottmann visualizes two conflicting forces in political geography: circulation (movement) and iconography (images or folk symbols).

[1] I wish to thank the Standard Oil Company of New Jersey for its assistance. Attendance at the 1958 "Jersey Roundtable" on the subject of nationalism helped to clarify the ideas developed in this paper; subsequent help from the company in supplying data and in freely discussing the issues involved were of crucial importance.

[2] Robert B. NcNee, "Functional Geography of the Firm, with an illustrative case study from the Petroleum Industry," *Economic Geography*, XXXIV(1958), pp. 321–37, and "Toward a more humanistic Economic Geography: the Geography of Enterprise," *Tijdschrift voor Economische en Sociale Geografie*, August, 1960, pp. 201–5.

[3] Richard Hartshorne, "The Functional Approach to Political Geography," *Annals*, Association of American Geographers, XL(1950), pp. 95–130. The terms "centrifugal" and "centripetal" were used by Hartshorne in accordance with a suggestion by C. C. Colby, "Centrifugal and Centripetal Forces in Urban Geograhpy," *Annals*, Association of American Geographers, Vol. 23 (1933), pp. 1–20.

[4] Jean Gottmann, *La Politique des Etats et leur Geographie* (Paris: Armand Colin, 1952), 228 pp.

[5] Stephen B. Jones, "A Unified Field Theory of Political Geography," *Annals*, Association of American Geographers, XLIV(1954), pp. 111–23.

Circulation encourages change while iconography is a more conservative force. Changes in circulation allow new forms of spatial organization; iconography tends to hallow and preserve the spatial organization developed in the past. This is significant for the functional geography of the firm because economic circulation is often organized by giant, functionally integrated, international companies, as in the oil business. For some, the organized space and circulation may include all or nearly all of the free world. But such global spatial circulation inevitably produces changes in the way of life of the various peoples linked together. Parochial or local iconography (cultural heterogeneity) may be inconsistent with the new "one-worldism" (economic homogeneity) of the global corporation. Resentment at this inconsistency may take the form of a highly emotional and even violent nationalism. Nationalistic fervor may lead to the erection of barriers to circulation which reduce the internal unity and coherence of the international company region, even though such barriers may not be in the immediate economic self-interest of such a nationalistic people. These barriers may even force the company region to contract in area, as in cases of state sponsorship of competing local companies, especially when such companies are given monopolies or near-monopolies by the state. The contraction may be violent, as in expropriations and other evidences of hostility toward international firms. These may result from something other than real or fancied economic or social injustices; they may derive from an inevitable conflict between an iconography which is local and an organization of economic circulation which is global. In this sense, the international corporation tends to create the very nationalistic feelings which it, for business reasons, so heartily deplores. The survival of international company regions may require more than the continued elimination of local economic or social injustices in royalties, wages, or housing, and continued adjustments to local cultural differences or nationalistic aspirations. Survival may be impossible without the growth of a global iconography to match the global organization of circulation by international firms. If so, iconography is a major limiting factor in the areal growth of company regions.

The Unified Field Theory

Jones has suggested that his unified field theory of political geography may be applicable also to analogous relationships in economic geography. The idea–area chain (idea–decision–movement–field–area) operates in both directions, with ideas ultimately creating areas and with existing areal organization "conditioning" future decisions, movements, fields, and areas. Thus a geologist's idea about the distribution of petroleum products might lead to a management decision to invest large sums in prospecting. This would lead to movement—negotiations for leases (usually with territorial states), intensive mapping, and exploratory drilling. Since such movements would be closely integrated areally, the pattern could be described as a "field." Eventually, the field, or at least a part of it, might be organized permanently as a corporate component of an international corporate regional system. If the parent company were an integrated one, participating in all phases of the oil industry, further decisions on refining, marketing, and transporting would now be necessary. These decisions would again lead to movement, new fields, and new corporate areas. But all decisions on prospecting and on the development of new oil pools are "conditioned" by the pre-existing areal organization. Company geography and the geography of territorial states are alike important in such conditioning. Thus decision on prospecting in a particular area would involve much more than a consideration of geology or of company need for additional crude oil. It would involve also a consideration of the company's areal marketing pattern and that of competitors, of the location and characteristics of company refineries, and of the areal transport system and its geographic flexibility. Further, it would involve careful weighing of the political and economic purposes of the specific territorial states involved. All of these decisions involve so many imponderables that mistakes are almost inevitable and survival requires constant reassessment of company geography.

Jones has characterized the international oil industry as "inherently dynamic, since new sources must be discovered."[6] There are many additional reasons for this dynamism, includ-

[6] *Ibid.*, p. 121.

ing the accelerating expansion of world energy demand, the corporate practice of re-investing profits which leads to massive international investments, and the highly dynamic leadership of most international petroleum companies. Today, one-third of the direct private foreign investments by Americans or American companies is in the petroleum industry. In 1956, such petroleum investment was at least 7 billion dollars.[7]

The Centrifugal–Centripetal Concept

The centrifugal–centripetal concept, as developed by Hartshorne for the analysis of territorial states, can also be applied to company regions, either domestic or international. Similar centrifugal forces (distance, physical barriers, human barriers, regional differences in economic development and economic relations, and sociocultural differences) tend to disrupt the areal unity and cohesion of the company region. For international company regions, the present chaotic system of conflicting sovereign territorial states is an additional major centrifugal force.

The international oil industry has exhibited a remarkable ability to overcome distance and physical barriers by innovations in transportation. The rise of oil to its present leading place in international trade is evidence of this ability. Yet distance and physical barriers remain important, as the Suez crisis indicated, though the precise meaning of such barriers is constantly changing with such innovations as the supertanker. The international oil companies often become involved in urban development as well as in providing their own transport and communications, since so much of the crude oil prospecting and crude oil production of recent decades has occurred in thinly populated and economically underdeveloped regions. Providing urban services is not only expensive, but it may lead also to "anti-paternalistic" resentment by company employees and local populations. Therefore, some oil companies seek to turn over such peripheral enterprises to local entrepreneurs or local government as soon as possible, even at substantial immediate financial loss to the company. Regional differences in economic relations find expression in such exchange problems as the "dollar gap" after World War II. Exchange controls and restrictions on the convertibility of currencies instituted by states may force a theoretically cohesive international petroleum company to operate, in practice, as a number of quite distinct areal units, as a "league" rather than an "empire." More often, such controls force the company to modify its internal pattern of circulation. Profits which cannot be exported directly from a given country may be exported in the form of new tankers, refinery equipment, or food supplies for company employees. To succeed, an international oil company must develop a degree of global cultural homogeneity, at least at top management levels. Top leadership must be trained and experienced not only in the technical and economic phases of the business, but also in the different culture areas of the world. Company decisions which are obvious and equitable to an American may appear otherwise to an Arab sheik or a Latin American politician.

But the most important centrifugal force for international company regions is the state, for the fundamental purposes of such companies may be in conflict with the fundamental purposes of the territorial states linked together by the company region. Nationalism, a powerful unifying force *within* territorial states is a disintegrating or centrifugal force for the world community as a whole. But whether nationalistic or not, a state's "main purposes are political. The values over which it has complete control are political values."[8] To achieve its ultimate political purposes, every territorial state must establish regional uniformity and coherence over a definite portion of the earth's surface. This must include a degree of regional uniformity and coherence in internal and external economic relations. For business enterprises, the fundamental purpose is *profit maximization*, at least according to classical economic theory. This does not necessarily mean immediate and short-run profits; in the modern, large-scale, corporation it is more likely to mean a concern for steady, long-range profit-making. However, the purposes of corporate enterprise may be more complex than classical economics suggests. Recent literature on the "managerial revolution" suggests that concern for the general stability

[7] Petroleum Industry Research Foundation, *United States and Oil Imports* (N.Y., 1958), p. 108.

[8] Hartshorne, *op. cit.*, p. 113.

and reputation of the company—social prestige, competitive gamesmanship (i.e., playing the "oil game"), and professional managerial morality or the "corporate conscience"—may be more important in motivating corporate management than mere profit maximization.[9] But in any case, there are likely to be substantial differences in purpose between international companies and territorial states. Such differences may lead to conflict, though this conflict is lessened when the state severely limits its control of economic relationships, as under the *laissez faire* philosophy of political economy, or when the state assumes control of all major economic enterprises, as in totalitarian states. In the mixed economy of the world today, some conflict between the territorial states and international business enterprise is generally inevitable. Although the territorial, economic, political, and strategic relations desired by a particular state may temporarily coincide with the business relations desired by an international company, such coincidence is often short-lived.

This is not to say that the coincidence of purposes of company and state is impossible or uncommon. Rather, it is to emphasize that such coincidence is normally an unstable and ephemeral phenomenon. Thus, throughout the history of petroleum various oil companies have used or attempted to use the power of various states to foster their growth while many states have used or attempted to use international oil companies to foster their policies. The international petroleum regions of today bear the marks of imperialism and the decline of imperialism. The British and Dutch governments were active in the early decades of the twentieth century in promoting companies led by their nationals and in excluding outside firms from their empires, mandated territories, and spheres of influence. This included oil firms as well as many others. The American government insisted on the "open door," i.e., the admission of American firms in foreign areas on the same bases that foreign firms were allowed to operate in the United States.[10] But overly close ties with specific national states can also be embarrassing to international business enterprise. Most American oil companies operating abroad today emphasize their purely business character and try to avoid becoming closely identified with specific aspects of American foreign policy. Company survivals in spite of recent revolutions in Venezuela, Iraq, and Indonesia seem to show the business wisdom of this policy.

It is in the centripetal forces integrating the region that the company region differs most markedly from a territorial state. The basic centripetal force for the state is the somewhat vague but nevertheless powerful feeling of belonging, the "state-idea." In modern states, the state-idea is linked most often with nationalism. This feeling of belonging is probably important in any human organization. Thus company periodicals tend to stress the "team spirit" and "family" aspect of the company. But although the feeling of belonging is never totally absent in company regions, it is probably weak in comparison with nationalism as a binding force. However, there is little concrete evidence on this point. Thus there is a real need for studies by social psychologists and social anthropologists to determine the relative strengths and weaknesses of company loyalties and national loyalties, especially in underdeveloped countries and in international firms. The major *economic* centripetal forces for company regions are legal ownership, the economies of scale, and, for industrial firms, vertical or functional integration.

Perhaps the most obvious centripetal force is legal ownership, a right granted in various forms by states but resting ultimately on a common acceptance of international law. In general, legal ownership implies that all parts of a corporate regional system will similarly obey the decisions of central company leadership and that, in democratically organized company regions, all parts will participate in much the same way in the formulation of such decisions. However, there are many limitations on this generalization, both in terms of corporate regional organization and in terms of the behavior of states. Partnerships, so common in the international oil industry, are a limiting factor because, in many cases, no single company holds a majority interest in the partnership, consortium, or "joint operation."

---

[9] Edward S. Mason, "The Apologetics of 'Managerialism'," *The Journal of Business of the University of Chicago*, XXXI(1958), pp. 1–11.

[10] Sen. Doc. No. 43, "Diplomatic Protection of American Oil Interests in Mesopotamia, Netherlands East Indies, and Mexico" (Washington, 1945).

Similarly, the rights and duties associated with ownership are continuously subject to re-definition by the state. Though a state may continue to respect a company's technical legal ownership of an oil property, it may radically alter the true meaning of ownership by changes in taxation, in labor legislation, or in the transferability of profits. The 1951 refinery agreement between India and the Standard-Vacuum Oil Company sought to resolve this problem, at least temporarily. It not only prohibited nationalization for at least 25 years, but also bound India not to apply import duties on crude oil or to require domestic prices for refined products below the import cost on such products.[11] An important limitation on the long-range effectiveness of legal ownership as a centripetal force is the wide geographic variation in the specific provisions of oil leases and the need to re-negotiate oil leases with the rapidly changing economic conditions of the oil industry. In Iran, the international oil companies now operate under "exclusive use" rights rather than the more usual concessional "ownership" rights. This fine distinction seems more significant legally than operationally. A special cause of confusion in legal ownership is the general lack of knowledge of mineral ownership law and of conservation laws based upon such law.[12] Perhaps it is because of the great geographic variations in the effectiveness of legal ownership as a centripetal force that each of the large international oil companies has developed a business diplomatic system, with "business diplomats" constantly shuttling from oil capital to oil capital, around the globe.

The economies resulting from large-scale operations are a major centripetal force. This is especially significant in industries such as petroleum in which a company characterized by vast, geographically dispersed operations is also characterized by very large-scale, semi-automatic producing and refining facilities. Economies of scale act as a deterrent to nationalization, especially in small or economically underdeveloped countries. For example, the nationalistic dispersal of refining would almost certainly mean much more expensive and less satisfactory petroleum products regardless of local managerial efficiency, except, perhaps, in a few large and economically advanced countries with a high-volume petroleum demand. Large-scale, centralized research may be considered a special case of the economies of scale.

The functional integration of the international petroleum industry is a major centripetal force, a counterforce to nationalistic fragmentation. The internationalism of oil is rooted in the wide divergence between the areas of petroleum production and the areas of consumption. Extensive functional integration within the territory of a single nation-state is uncommon because few states have both large reserves and large consumption. The United States, Canada, and the U.S.S.R. are conspicuous exceptions. As Hartshorne has succinctly stated, the crucial significance of trade for a territorial state "depends on the possibility of alternatives, of finding other sources of needed supplies or other markets for products which must be sold to maintain the national economy."[13] Thus, in the Iranian nationalization crisis, the deficiency of technical skills was probably not a major problem for Iran, despite popular opinion to the contrary. Probably the lack of tankers, the inability to open large-scale markets quickly, and the difficulties in obtaining repair parts were much more important problems for Iran. Apparently the Mossadegh government did not consider such problems seriously before nationalization.[14] Similarly, the possibility of alternatives is important for company regions. That is, an international company with a regional system of markets, refineries, and crude petroleum production widely dispersed in many states is least vulnerable to nationalistic pressures.

CENTRIFUGAL–CENTRIPETAL ANALYSIS OF THE STANDARD OIL COMPANY OF NEW JERSEY

The following limited analysis of the Standard Oil Company of New Jersey is intended as a concrete illustration of the concept of centrifugal–centripetal forces in international petroleum company regions. It is necessarily

[11] L. M. Fanning, *Foreign Oil and the Free World* (N.Y.: McGraw–Hill, 1954), p. 204.

[12] N. J. Campbell, Jr., "Principles of Mineral Ownership in the Civil and Common Law System," *Tulane Law Review*, 1957.

[13] Hartshorne, *op. cit.*, p. 123.

[14] Personal communication from Dr. Jahangir Amuzegar, formerly employed by the Iranian government and presently Assistant Professor of Economics at Michigan State University.

limited and partial because of the great size, vast scope, and extensive history of the Jersey system. Thus it is more suggestive than definitive.

Jersey Standard is the giant of an industry noted for great corporate size. It vies with General Motors for top honors in corporate size among all American companies. It is the largest single petroleum company in the world, although its major competitor, "the Group" (Royal Dutch–Shell), is close behind and other petroleum systems such as Gulf, Texas, Standard Oil of California, Socony Mobil, British Petroleum–Sinclair, and the Russian State system are also very large. Jersey Standard supplies about 16 percent of the free world's supply of petroleum products. Its total *gross* revenue in 1957 was nearly 8 billion dollars; its property and equipment were worth at least 5 billion dollars. Size alone would warrant careful scrutiny of Jersey's functional geopraphy. But Jersey is also outstanding in many other respects. It has been a pioneer in the petroleum industry since its founding over 75 years ago in 1882. It has tried to maintain its leadership not only by constant research and experimentation in the conventional phases of the petroleum industry but also in more peripheral industries, such as petro-chemicals. It is perhaps the most international company in an industry noted for internationalism. About three-fourths of Jersey's crude oil and two-thirds of its refined products are produced by affiliates outside the United States; over 60 percent of the markets of all affiliates are abroad, if one calculates markets by volume. Nearly three-fourths of Jersey's income derives from these foreign operations.[15]

The Jersey Standard Corporate Regional System

The corporate region of Jersey Standard is a delicately balanced managerial and investment system, reflecting the importance of both centrifugal and centripetal forces in a global company region (Fig. 1 and Table 1). According to conventional business terminology, Jersey Standard is classified as a holding company. That is, it holds shares in operating companies and in other holding companies which, in turn, hold shares in operating companies. But it is a rather special type of holding company because its functions extend far beyond the mere financial supervision of its global investments. These functions are centripetal, integrating the global petroleum region. Jersey Standard's regional system operates on the federal principle, with management highly decentralized. The parent company, Jersey Standard, leaves the direct operating functions (exploration, crude oil production, transporation, refining, marketing, and research) to the component companies of the system while retaining for itself only those functions which clearly require global centralization. Jersey Standard always has been organized federally, though decentralization has been expanded gradually through the years. The original Standard Oil Company was conceived by John D. Rockefeller as a federal system, modeled after the then decentralized federal system of the United States. Through the years, as the company has grown and spread over the earth, it has become more and more decentralized. Contrarily, in the same period, the United States has become more centralized, with the component states relinquishing many governmental functions to the national government.

The component companies of the Jersey Standard global system vary widely in the size and nature of their operational areas. In Table 1 these companies have been classified broadly as either *zonal* or *global* companies. Only the more important of the zonal companies are indicated separately in Figure 1. The larger and more important zonal companies are at least partially integrated, i.e., they combine two or more of the basic petroleum functions. The lesser firms are more specialized, including many firms specializing in marketing only. The zonal firms generally confine their activities to definite and restricted portions of the earth's surface, although a few, such as Standard–Vacuum, have very large operational areas. The operational areas are defined generally in terms of the boundaries of territorial states, i.e., they are confined to individual territorial states or internal governmental units thereof, or to regional groupings of territorial states.

The zonal firms are, at least in part, an

[15] Standard Oil Company of New Jersey, "75th Anniversary of Jersey Standard," *The Lamp*, 1957, p. 52.

TABLE 1.—MAJOR CORPORATE DIVISIONS OF JERSEY STANDARD SYSTEM IN 1958[1]

Corporate unit	Percent owned	Functional integration[2]	Special comment
		Global Companies	
Esso Export Co.	100	M	World-wide wholesaling; coordination of bunkering and aviation sales
Esso Shipping Co.	100	Tm	
Esso Export, Ltd.	100	Mw	Wholesaling, especially in sterling area
Panama Transport Co.	100	Tm	
Standard Oil International, Ltd.	100	Mw	Wholesaling, especially Middle East crude oil
Mediterranean Standard Oil Co.	100	Mw	Wholesaling crude and refined oil from Middle East
Enjay Co., Inc.	100	M	Marketing chemical products
Gilbert & Barker Manufacturing Co.	100	——	Oil burners; service station equipment
Ethyl Corp.	50	——	Anti-knock compound made and sold
Esso Research and Engineering Co.	100	——	Research and development
Jersey Production Research Co.	100	——	Production research
		Zonal Companies	
Esso Stand. Oil Co.	100	T, R, M	U.S., Atlantic Seaboard
Carter Oil Co.	100	P, T, R, M	U.S., Pacific Northwest
Humble Oil and Refining Co.	98	P, T, R, M	U.S., Central and South Central
Interstate Oil Pipeline Co.	100	Tp	Humble-area; minor in Carter area
Other pipeline companies (four)	6–49	Tp	Imperial, Carter, Esso areas
Imperial Oil, Ltd.	69	P, T, R, M	Canada
Lago Oil & Transport Co., Ltd.	100	R	Aruba (Netherlands Antilles)
Creole Pet. Corp.	95	P, T, R, M	Venezuela
International Pet. Co., Ltd.	97	P, T, R, M	Peru, Ecuador, Colombia, Venezuela (minor)
Esso Stand. Oil, S.A.	100	R, M	Central America, Caribbean, Guianas
Esso S.A. Pet. Arg.	100	P, T, R, M	Argentina
Arabian–American Oil Co. and Tapline	30	P, T, R (M)	Marketing limited to Saudi Arabia
Iraq Pet. Co., Ltd. and assoc. co.	11.9	P, T, R (M)	Marketing limited to Iraq
Esso Trading Co. of Iran	100	P, T, R, M	About 7 percent share in Consortium
Standard–Vacuum Co.	50	P, T, R, M	Far East, Australasia, Africa
Esso Standard, S.A.F.	63	P, T, R, M	Continental France
Esso Pet. Co., Ltd.	100	T, R, M	United Kingdom and Eire
Esso A. G.	100	R, M	West Germany
Gewerkschaft Brigitta	50	P	West Germany
Stanic-Industria Petrol, S.p.A.	50	R	Italy
Esso Stand. Ital., S.p.A.	100	M	Italy
Esso (Belgium)	100	R, M	Belgium
Esso Standard Refin.	100	R	Belgium
Esso-Nederland N.V.	100	M	Netherlands
N.V. Ned. Aardoli Mij.	50	P	Netherlands
Dansk Esso A/S	82	R, M	Denmark
Ak. Norske Esso	56	R, M	Norway
Other companies	100	M	Largely marketing companies operating in Latin America, Europe, Mediterranean, Nigeria, and Congo

[1] Data from stockholder reports and other company publications. Special comments and classification of the companies are by the author.

[2] Functional integration symbols as follows: P: crude oil production; T: transportation; Tp: pipeline transport only; Tm: marine transport only; R: refining; M: marketing; Mw: wholesale marketing only.

adjustment to the many centrifugal forces affecting the petroleum company region. Among the zonal companies partial ownership by Jersey Standard is common whereas nearly all of the global firms are 100 percent owned by Jersey Standard. Ownership may be shared with the other international oil "giants," as in the Middle East, to diversify investments and share the risks. In other cases, participation by local citizens or firms promotes good will

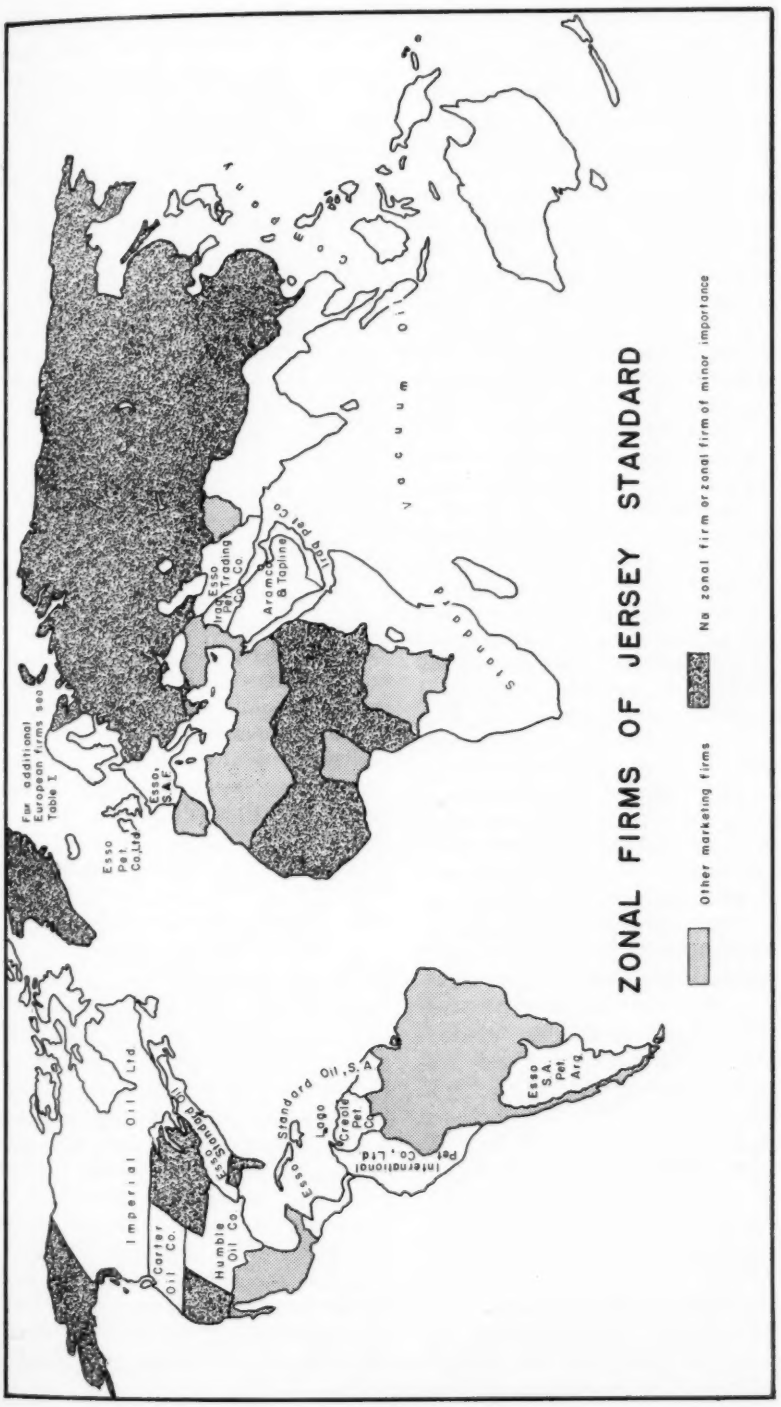

ZONAL FIRMS OF JERSEY STANDARD
Other marketing firms
No zonal firm or zonal firm of minor importance
Vacuum Oil Co
Pipeline
Arabian Pet Co
Aramco & Tapline
Iraq Esso Pet Trading Co
Esso Pet Co Ltd
Esso SAF
For additional European firms see Table I
Imperial Oil Ltd
Esso Standard Oil S A
Carter Oil Co.
Humble Oil Co
Standard Oil S A
Lago
Creole Pet Co
International Pet Co Ltd
Esso SA Pet Arg

FIGURE 1

as well as reducing the investment required. In one case, Jersey shares ownership equally with a state. This is Stanic–Industria Petrolifera, S.p.A., jointly owned with the Ente Nazionale Idrocarburi, an Italian government firm. This arrangement was initiated by the Italian state, rather than by Jersey Standard.

Zonal firms have many operational advantages. Such a firm can conduct easily its business in the language of a particular state and within its monetary and legal framework. A zonal firm can employ nationals primarily, even including nationals in top management positions to ensure company decision and practice appropriate to the culture, legal systems, and state regulation of the area. However, Standard–Vacuum operates in an area which is so heterogeneous culturally and so fluid in its legal and political framework that the company has many problems in adjusting to local differences while striving to maintain its basic unity.

The *global* companies of the Jersey system differ from the zonal firms in many ways; they are instruments of Jersey Standard in carrying out its centripetal functions. For example, the British government limited the importing of crude oil or refined products after World War II in order to conserve foreign exchange. Though adversely affected by this policy, the Esso Petroleum Company, Ltd., a zonal firm, was powerless by itself to change the situation. Later, petroleum rationing was lifted in Great Britain as the result of a complex system of dollar conservation worked out by the global organization of Jersey Standard and other international oil companies. Some of the global companies, such as some shipping and wholesaling firms, operate over much of the free world. On the other hand, the Esso Research and Engineering Company and the Jersey Production Research Company operate largely in the United States. But they have been classed as global companies because their findings are applied by Jersey affiliates throughout the Jersey Standard global region. When the discoveries of these companies are patentable, they help to integrate the Jersey global region, at least for all territorial state-areas which recognize international patent agreements.

The parent company, Jersey Standard, has reserved for itself the most difficult managerial responsibilities. The magnitude of these responsibilities is inadequatelv indicated by the size of its employment, only about 1,400 persons. This is less than one percent of the total employment in the whole company region. The general goal of the company is world-wide uniformity in the continuous availability of refined petroleum products at prices competitive with non-Jersey sources of petroleum products and with alternative energy sources, such as coal or water power. This requires the coordination and integration of geographically dispersed crude-oil production zones, refineries, and petroleum markets, and the elaborate transport network which links them together. It also requires careful planning for balanced growth in the development and use of petroleum resources, i.e., parallelism among discovery and exploitation, the development of new kinds of petroleum products, and the expansion of petroleum markets. In short, Jersey Standard's basic function is to provide strategic long-range planning for unity and coherence in the global region.

A special aspect of Jersey's centripetal function is to act as a counter-force to the fragmenting influence of the national state, a major centrifugal force. In the short range, this can be accomplished by tactical changes in the global pattern of functional integration, i.e., by temporarily altering global petroleum circulation. In the long range, the centripetal goal can be achieved by basic research in petroleum science and by directing investment to those states where centrifugal forces appear most attractive to the whole system.

A possible limiting factor in exercising the centripetal function is the company's American locus and American heritage. Despite its internationalism, company headquarters are in the United States (30 Rockefeller Plaza, New York); top management and most of its investors are American citizens. The company's version of the modern managerial business creed, of "business with a heart," is a product of the American enviroment. In short, Jersey Standard is an American company, subject to many pressures from American society and from the American state. Its world-wide system is viewed by friend and foe alike as a significant part of the total pattern of American economic leadership throughout the free world.

Crude Oil Production, Reserves, and Exploration

The current geographic pattern of crude oil production, reserves, and exploration within the Jersey regional system reflects the dynamic, expansionist policy of the company as well as the peculiarities of geologic distribution of crude petroleum around the world. It is also a record of the past successes and failures of Jersey Standard in counteracting the centrifugal forces disrupting the region. Generally, the company has been less successful in overcoming the fragmenting influence of sovereign states than in overcoming such centrifugal forces as distance, physical barriers, or regional disparities in population distribution and economic development. The most dramatic past failures in relations with territorial states include expropriations in Bolivia (1937), Mexico (1938), and Eastern Europe (first by the Nazis, later by the Communists). Equally important, though less dramatic, has been the slow decline in oil production in such countries as Argentina because of the inability to win new concessions or exploratory rights. (New agreements signed in 1958 between Jersey Standard and Argentina allow exploration in the oil-rich province of Neuquen.) A nontraumatic but limiting factor in the growth of the system has been the inability to win any producing foothold at all in such countries as Brazil, though Jersey has long been an important marketer in Brazil. However, the areal expansion of the company indicates that centripetal forces have been stronger than centrifugal forces over the last 75 years.

Currently, Jersey Standard is rather vulnerable to nationalistic pressures, though less so than such other petroleum firms as Gulf Oil or British Petroleum. Affiliates of the Jersey Standard family produce petroleum in many widely scattered parts of the free world (Fig. 2).[16] However, most of the petroleum production is remarkably concentrated in only a few areas. Creole of Venezuela alone produces about 48 percent of the total while the Persian Gulf companies account for about 20 percent, the U.S. about 20 percent, Canada about 4 percent, and all other countries only about 8 percent. This vulnerability is even more marked in terms of proved reserves, since the Persian Gulf countries have about 50 percent, Venezuela about 31 percent, the U.S. about 14 percent, and all others only about 5 percent, according to figures available for 1958. Subsequent discoveries, such as those in Libya, have somewhat altered this pattern, but only to a minor degree.

In terms of present-day production, Creole's crude oil is as pivotal to the economy of Venezuela as it is to the Jersey Standard global system. In this mutual dependence there is a solid foundation for continued cooperation. Internationally, both Venezuela and Creole have the same broad interest in promoting freer world trade in petroleum and its products. The diplomatic voice of Venezuela is supplemented by the "business diplomatic" voice of Jersey Standard in Washington, London, and other political capitals around the world. The domestic relationships between Creole and Venezuela were considered classic examples of cooperation between foreign private enterprise and the state in the Paley report. Creole (Jersey Standard) hopes to stay in Venezuela indefinitely.[17] Therefore the company goes far beyond mere compliance with the complicated Venezuelan laws on conservation, profit-sharing, employment practices, and the like. It tries to identify itself with the people of the country by domestic purchase of supplies, financing of certain public projects, aiding higher education through the Creole Foundation, integrating with the community by eliminating the old isolation of the oil camps, and offering Creole shares on the Venezuelan stock exchange.

The Persian Gulf countries are a more difficult problem. Just how Jersey can accommodate itself to the rising nationalism of the area with safety is not yet clear. Demands for a greater share in profits can be met more easily than the demands for state participation in integrated operations, including tanker fleets,

[16] Crude oil production in barrels per day was as follows in 1957: Venezuela: 1,249,600; Humble area: 420,000; Middle East: 416,900 (Saudi Arabia: 297,600; Iraq: 52,600; Iran: 50,100; Quatar: 16,600); Canada: 109,700; Carter area: 80,800; Peru: 35,300; Indonesia: 32,500; France: 24,500; Netherlands: 14,600; West Germany: 2,600; Argentina: 800. Data supplied by the Standard Oil Company of New Jersey.

[17] W. C. Taylor and J. Lindeman, *The Creole Petroleum Corporation in Venezuela*, Fourth Case Study of United States Business Performance Abroad, National Planning Association, Washington, 1955, 103 pp.

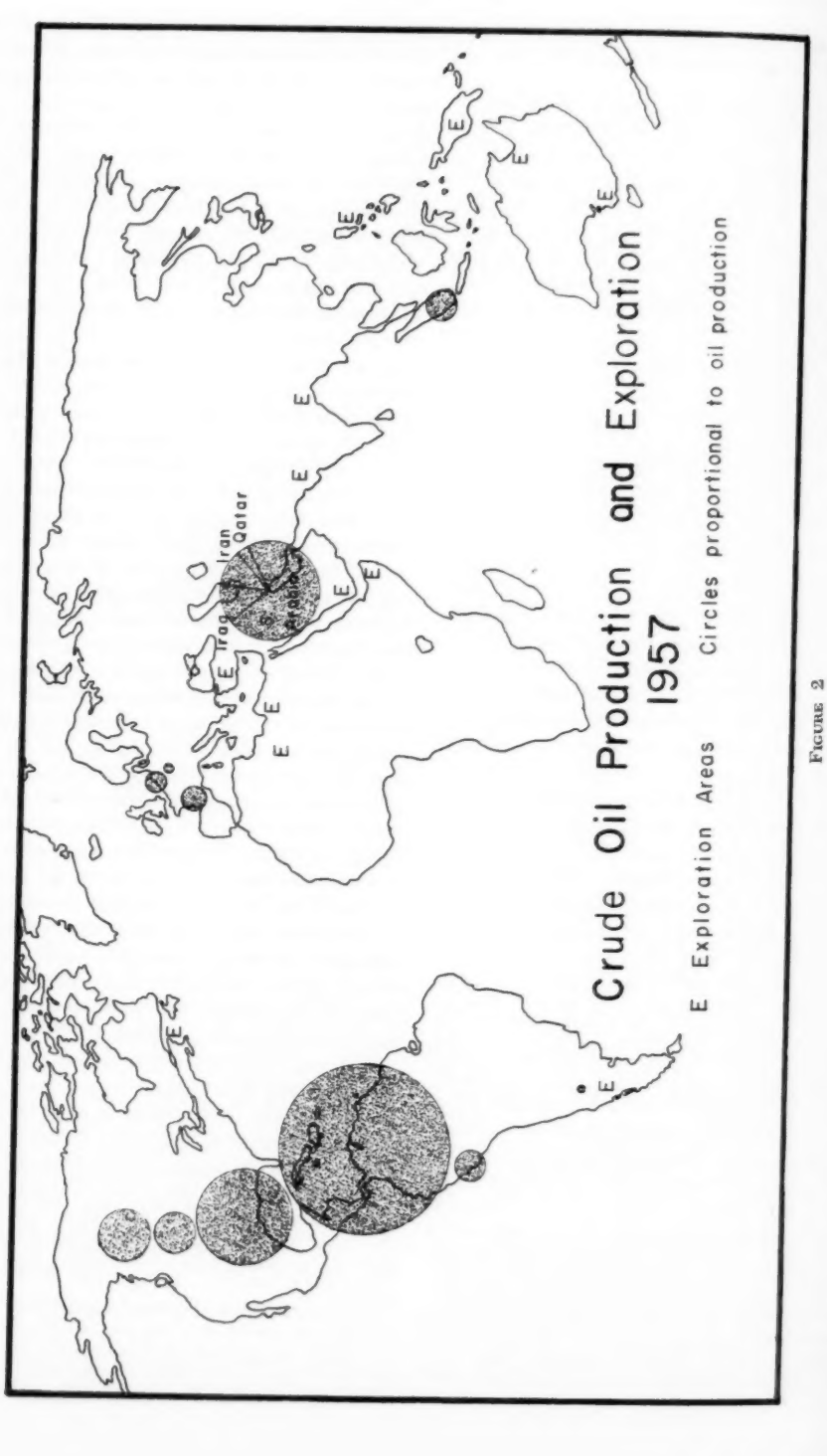

Crude Oil Production and Exploration
1957
E Exploration Areas
Circles proportional to oil production
Iran
Qatar
Iraq
Saudi Arabia

FIGURE 2

refining, and global marketing. The participation of territorial states directly in global management might seriously hamper the centripetal function of Jersey Standard.

Today there is a relative abundance of crude oil reserves and production within the Jersey Standard system and for the oil industry as a whole because of the phenomenal discoveries of the Persian Gulf area. Nevertheless, affiliates of the Jersey system are very actively searching for new deposits in many widely scattered areas of the world. A large organization such as Jersey Standard must plan many years in advance. Over the past 45 years, for example, world petroleum demand has increased by 15 million barrels per day and it is estimated that in the next 15 years it will grow by another 15 million barrels per day. Also, Jersey could be more resistant to nationalistic pressures if oil production and reserves were more widely scattered in more territorial states. However, the exploratory effort is restricted by the hostility and ambivalence of some territorial states to foreign oil companies. Although Indonesia has recognized the legality of the oil leases made with the Dutch Indies, it has been slow to establish a clear policy for new oil leases.[18] The east slope of the Andes south of Venezuela is an area of very promising oil potential, but Brazil, Bolivia, and Argentina historically have restricted severely the exploratory operations of foreign firms. The relaxation now evident in Bolivia and Argentina may, in time, spread to Brazil but at present exploration in Brazil is monopolized by a small government firm, Petro-Braz.

Jersey Standard Marketing and Refining

The Jersey Standard companies are major marketers in most of the free world (Fig. 3). The core marketing area is the North Atlantic littoral. Marketing is rather minor, however, in several Latin and Latin American states, such as Spain, Mexico, and Bolivia. The only important marketing foothold in West Africa is a minor firm operating in and near Dakar. Jersey has not yet penetrated significantly the markets of the United States Middle West, although it has recently acquired several marketing outlets in the Chicago area (Pate Co.). Most of the American Southwest and much of the Southeast are largely untouched by Jersey marketing. In short, Jersey has developed international marketing more extensively than domestic marketing. Though it has an important share in most of the major free world markets outside the United States, it has not yet developed a truly national domestic marketing program.[19]

The refining core of the Jersey world region is the North Atlantic littoral, with the most pronounced clustering in three areas: (1) Gulf Coast–Caribbean, (2) northeastern North America, and (3) Western Europe. The Gulf Coast–Caribbean cluster has more than half of the total capacity; it reflects primarily source attractions. This cluster is the key to the refining phase of Jersey Standard's centripetal function. The great surpluses of refined products from this area can be directed to almost any part of the world, either to supply markets totally or to supplement inadequate local refining. The northeastern North American cluster and the Western European cluster respectively have one-eighth and one-sixth of the total capacity; they are somewhat less flexible in operations because they are primarily oriented to large local markets. These clusters have been influenced to an indeterminate degree by state pressure or nationalistic forces.[20] Outside these three clusters, the refineries are more widely scattered, without strong clustering. These widely scattered refineries are primarily source refineries and secondarily market refineries. A few are nodally located. It is evident that the importance of the various locational factors differs very significantly in relation to company functional geography, since combination refineries and market refineries are much less important in the Jersey world region than in the Socony Mobil world region.[21]

Although there are more than 40 refineries

[18] Center for International Studies, Massachusetts Institute of Technology, *Stanvac in Indonesia*, Sixth Case Study of United States Business Performance Abroad, National Planning Association, Washington, 1957, 118 pp.

[19] The consolidation of all domestic firms under the Humble Oil Company, promulgated in 1960, may presage the development of a truly nationwide marketing system.

[20] Alexander Melamid, "Geographical Distribution of Petroleum Refining Capacity, a Study of the European Refining Program," *Economic Geography*, XXXI(1955), pp. 168–78.

[21] McNee, *op. cit.*

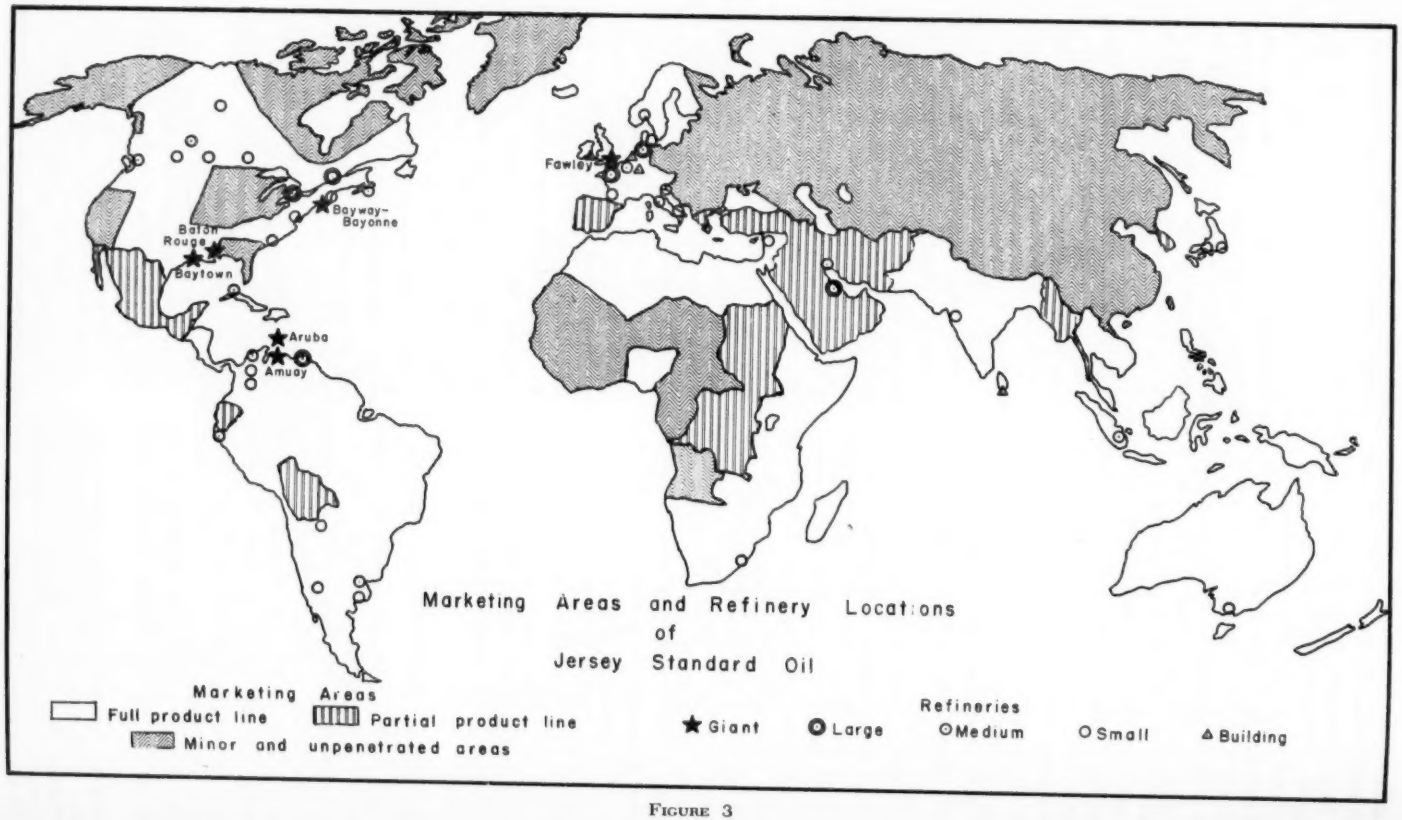

Fawley
Bayway-
Bayonne
Baton
Rouge
Baytown
Aruba
Amuay
Marketing Areas and Refinery Locations
of
Jersey Standard Oil
Marketing Areas
Full product line
Partial product line
Minor and unpenetrated areas
Refineries
Giant
Large
Medium
Small
Building

FIGURE 3

in the global system, the economies of large-scale operations have tended to concentrate most of the refining capacity in only a few giant refineries. In Figure 3, the refineries have been classified as *giant* (over 100,000 barrels per day capacity), *large* (50,000 to 100,000), *medium* (25,000 to 50,000), and *small* (under 25,000), according to crude capacity[22] and percentage of ownership.[23] The six giant refineries (Aruba, 440,000; Amuay 328,000; Baton Rouge 290,000; Baytown 282,000; Fawley 228,000; and Bayway–Bayonne 145,000 barrels per day) have about two-thirds of total capacity. Aruba is currently the largest single refinery in the world; it has about one-sixth of total Jersey capacity. Since Aruba is nodally located on a Dutch island, though near its normal crude oil source in Venezuela, its strategic importance for the whole system is difficult to exaggerate. However, Aruba's unique importance, though not its capacity, will no doubt decline in the future. Today the combined crude capacity of Creole's two Venezuelan refineries, Carpito and Amuay, is nearly as great as that of Aruba. Current and prospective concession agreements in Venezuela will ensure an important refining industry in Venezuela. Significantly, even the specific site of the Amuay refinery was selected after consultation with the Venezuelan authorities.[24]

An international company such as Jersey Standard cannot escape the influence of territorial states, but it can and has cooperated with specific states in developing a global petroleum region in which the centripetal and centrifugal forces are delicately balanced. Continued cooperation between Venezuela and Jersey Standard is very important to the continued growth and coherence of the world-wide region; a breakdown in such cooperation might require sweeping changes in many parts of Jersey's global region. Cooperation has political as well as economic advantages for Venezuela. Through its crucial role in the Jersey Standard region, it can exert significant influence on political events throughout the globe.

[22] Data adapted from Bureau of Mines, "Capacity of Petroleum Refineries and Cracking Plants in the United States: Jan. 1, 1958," special release, 1958, and *World Petroleum*, 28th annual Refining Review, July 15, 1958; cracking capacity is concentrated at the larger refineries, making them proportionally more important than crude capacity alone indicates; it should be noted that the map indicates *capacity* rather than actual refining. Therefore, it is not fully comparable with the map of crude oil production which indicates *production* rather than *capacity*.

[23] For example, Abadan is one of the world's largest refineries but it is relatively unimportant for the Jersey system because Jersey's interest in the Iranian Consortium is small.

[24] Taylor and Lindeman, *op. cit.*, pp. 17–19.

SUMMARY AND CONCLUSIONS

Company regions are analogous, to a degree, to territorial states. Consequently, concepts from political geography may be used to illuminate the functional geography of the firm. The centrifugal–centripetal concept is especially applicable to the analysis of international firms, such as those of the petroleum industry. The unifying, or centripetal, forces include company loyalty, a force somewhat analogous to the state-idea. The most important economic centripetal forces are legal ownership, economies of scale, and functional integration. Company regions are affected by the same disintegrating or centrifugal forces operating in territorial states. In addition, the national territorial state is a major centrifugal force for international company regions. Conflict between international company regions and territorial states is almost inevitable because of the differences in purpose for which they are organized.

The Standard Oil Company of New Jersey is organized on the federal principle, allowing local adjustments to various centrifugal forces, yet retaining global unity by central coordination. The current crude oil production, marketing, and refining patterns of the company indicate that economic centrifugal–centripetal forces have been dominant in company growth. However, the geography of territorial sovereignty has modified and conditioned the pattern locally. The Jersey Standard global region is very dependent upon cooperation with the Venezuelan state, since the Venezuelan area is the production and refining core of the whole region. The company cannot hope to escape the centrifugal forces of the national territorial state; it can only hope to balance centrifugal forces with centripetal forces and thus maintain the unity of its global region.

This study has many implications for future studies in the functional geography of firms. It emphasizes the role of centrifugal–centri-

petal forces in only one type of company region, that of the *international* firm. This type of analysis could be applied also to company regions entirely contained within the boundaries of single territorial states. If this were done, it should be possible to determine the differences in functional geography between domestic and international firms with more precision than is now possible. Furthermore, this study has been concerned only with the *petroleum* industry. If centripetal–centrifugal analysis were applied to such firms as the Aluminum Company of America or General Motors one might find that the major centripetal–centrifugal forces were of a quite different nature. The comparative study of several international firms in several industries might allow the development of generalizations applicable to all international company regions.

This study also has wider implications for the study of the functional geography of institutions or the functional geography of organizations. The study implies that the unification or space-organization achieved by one social instrument or institution may tend to disintegrate the space-organization being developed by another institution, e.g., the conflict and compromise between territorial state spatial organization and international corporate spatial organization. But there are also cases in which one form of spatial organization tends to strenghten and preserve another. The nature of such conflicts or strengthenings could be clarified by an expansion of these studies to regional analysis of labor unions, religious bodies, and other major social institutions. Through such studies, geographers could make major contributions to the understanding of the spatial framework of society, from local and provincial levels to the whole earth.

aphy of in-
phy of or-
t the unifi-
ved by one
ay tend to
being de-
, the con-
torial state
nal corpo-
e are also
organiza-
e another.
gthenings
of these
or unions,
social in-
ographers
he under-
f society,
he whole